

| | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|---------------------------------------------------------------------|
| U.S. Department of Commerce Patent and Trademark Office | | Attorney's Docket No. 4894 |
| TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371 | | U.S. Application No. (if known, see 37 CFR 1.5) 09/623326 |
| INTERNATIONAL APPLICATION NO. PCT/EP99/01674 | INTERNATIONAL FILING DATE 15 March 1999 | PRIORITY DATE CLAIMED 13 March 1998 |
| TITLE OF INVENTION POLYMERASE CHIMERAS | | |
| APPLICANT(S) FOR DO/EO/US FREY, VILLBRANDT, SCHOMBURG, SOBEK AND ANKENBAUER | | |
| Applicants herewith submit to the United States Designated/Elected Office (DO/EO/US) the following items and other information | | |
| <ol style="list-style-type: none"> 1. <input checked="" type="checkbox"/> This is a FIRST submission of items concerning a filing under 35 U.S.C.371 2. <input type="checkbox"/> This is a SECOND or SUBSEQUENT submission of items concerning a filing under 315 U.S.C. 371 3. <input checked="" type="checkbox"/> This express request to begin national examination procedures (35 U.S.C. 371(f) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1). 4. <input checked="" type="checkbox"/> A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date. 5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2)) <ol style="list-style-type: none"> a. <input checked="" type="checkbox"/> is transmitted herewith (required only if not transmitted by the International Bureau). b. <input type="checkbox"/> has been transmitted by the International Bureau c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US) 6. <input checked="" type="checkbox"/> A translation of the International Application into English (35 U.S.C. 371(c)(2)). 7. <input checked="" type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)) <ol style="list-style-type: none"> a. <input checked="" type="checkbox"/> are transmitted herewith (required only if not transmitted by the International Bureau). b. <input type="checkbox"/> have been transmitted by the International Bureau. c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired. d. <input type="checkbox"/> have not been made and will not be made. 8. <input checked="" type="checkbox"/> A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)). 9. <input checked="" type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)). [unexecuted] 10. <input type="checkbox"/> A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)). <p>Items 11 to 16 below concern other document(s) or information included:</p> <ol style="list-style-type: none"> 11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98. 12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included. 13. <input checked="" type="checkbox"/> A FIRST preliminary amendment. <input type="checkbox"/> A SECOND or SUBSEQUENT preliminary amendment. 14. <input type="checkbox"/> A substitute specification. 15. <input type="checkbox"/> A change of power of attorney and/or address letter. 16. Other items or information: | | |

35,750
REGISTRATION NUMBER

09/623326
534 Rec'd PCT/PTC 30 AUG2000
PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application
based on International Application No. PCT/EP99/01674
Filed March 15, 1999
Inventor(s) FREY *et al.*

For: **POLYMERASE CHIMERAS**

Attorney Docket No. 4894

PRELIMINARY AMENDMENT

Assistant Commissioner for Patents
BOX PCT
Washington, D.C. 20231

Alameda, CA 94501
Date: August 30, 2000

Sir:

Prior to examining the above-referenced application as entering the National Stage under 35 U.S.C. §371, please consider the following amendments and remarks.

IN THE CLAIMS

At page 1, line 1 of the Amendment of claims under PCT Article 19, please delete "Claims" and insert therefor --WHAT IS CLAIMED IS--.

Please amend the claims as follows:

1. (Amended) A polymerase [Polymerase] chimera comprising [composed of] functional amino acid fragments of at least two different polymerases, wherein the functional amino acid fragments are active in the polymerase chimera, a [and the] domain having polymerase activity is derived from the first polymerase and a [the] domain having 3'- 5' exonuclease activity is derived from the second polymerase, and wherein the amino acid sequence of the polymerase chimera essentially corresponds to SEQ ID NO: 8.
2. (Amended) A polymerase [Polymerase] chimera comprising [composed of] functional amino acid fragments of at least two different polymerases, wherein the functional amino acid fragments are active in the polymerase chimera, a [and the] domain having polymerase activity is derived from the first polymerase and a [the] domain having 3'- 5' exonuclease activity is derived from the second polymerase, and wherein the amino acid sequence of the polymerase chimera essentially corresponds to SEQ ID NO: 10.
3. (Amended) A polymerase [Polymerase] chimera comprising [composed of] functional amino acid fragments of at least two different polymerases, wherein the functional amino acid fragments are active in the polymerase chimera, a [and the] domain having polymerase activity is derived from the first polymerase and a [the] domain having 3'- 5' exonuclease activity is derived from the second polymerase, and wherein the amino acid sequence of the polymerase chimera essentially corresponds to SEQ ID NO: 12.
4. (Amended) The polymerase [Polymerase] chimera as claimed in one of the claims 1-3, wherein the chimera additionally has reverse transcriptase [RT] activity.
5. (Amended) The polymerase [Polymerase] chimera of claim 4 [as claimed in one of the claims 1-4], wherein histine tags have been incorporated into the amino acid sequence of the chimera.
6. (Amended) A nucleic acid that encodes the [DNA sequence of a] polymerase chimera as claimed in claim 1, 2 or 3 [one of the claims 1-5].

7. (Amended) A nucleic acid that encodes [DNA sequence of] a polymerase chimera comprising the sequence of SEQ ID NO. 2 [according to SEQ ID NO. 2].
8. (Amended) A nucleic acid that encodes [DNA sequence of] a polymerase chimera comprising the sequence of SEQ ID NO. 4 [according to SEQ ID NO. 4].
9. (Amended) A nucleic acid that encodes [DNA sequence of] a polymerase chimera comprising the sequence of SEQ ID NO. 6 [according to SEQ ID NO. 6].
10. (Amended) A vector comprising the nucleic acid [Vector containing a DNA sequence] as claimed in claim 6 [claims 6-9].
11. (Amended) A host [Transformed] cell which has been transformed with [contains] the vector as claimed in claim 10.
12. (Amended) A process [Process] for the production of the polymerase chimeras as claimed in one of the claims 1-3 [1-5], wherein the process comprises the following steps:
 - (a) designing variants with the aid of amino acid sequence alignments, of three dimensional [3D] models or with the aid of experimentally determined three dimensional [3D] structures;
 - (b) production of domain exchange variants by genetic engineering;
 - (c) ligating [the] DNA fragments that encode the variants into starting vectors;
 - (d) expression of the chimeras in a host which has been [was] transformed by vectors carrying the DNA fragments; and

- (e) purifying the expressed polymerase chimeras.

13. (Amended) A method for using [Use of] the polymerase chimeras as claimed in one of the claims 1-3 [1-5] for PCR.

14. (Amended) A method for using [Use of] the polymerase chimeras as claimed in one of the claims 1-3 for sequencing [1-5 to sequence] DNA fragments.

15. (Amended) A method for using [Use of] the polymerase chimeras as claimed in one of the claims 1-3 [1-5] for RT-PCR starting with an RNA template.

16. (Amended) A kit comprising [Kit containing] a polymerase chimera as claimed in one of the claims 1-3 [1-5].

REMARKS

Applicants have amended the claims to comply with the U.S. patent practice in matters of form and to remove multiple dependency in certain claims. After entry of this Amendment, claims 1-16 are pending in this application. The amendments are fully supported by the specification, and do not introduce new matter. Entry of this Amendment is respectfully requested.

The total filing fee on the Transmittal Letter To The United States
Designated/Elected Office (DO/EO/US) Concerning A Filing Under 35 U.S.C. §371 is
calculated on the basis of this Amendment.

Respectfully submitted:

By: 

Victor K. Lee, Ph.D.
Agent for Applicants
Reg. No. 35,750
1145 Atlantic Avenue
Alameda, CA 94501
Telephone (510) 814-2966
Telefax (510) 814-2973

Polymerase chimeras

The invention concerns polymerase chimeras which are composed of amino acid fragments representing domains and which combine properties of naturally occurring polymerases that are advantageous with regard to a particular application. It has surprisingly turned out that the domains from the various enzymes are active in the chimeras and exhibit a cooperative behaviour. The present invention especially concerns those polymerase chimeras in which the domains having polymerase activity and domains having 3'-5' exonuclease activity are derived from different enzymes. Such chimeras can also have RT activity. In addition the present invention concerns a process for the production of the chimeras according to the invention and the use of these chimeras for the synthesis of nucleic acids e.g. during a polymerase chain reaction. Moreover the present invention concerns a kit which contains the polymerase chimeras according to the invention.

According to Braithwaite, D.K. and Ito, J. (1993) Nucl. Acids Res. 21, 787-802 DNA polymerases are divided according to the correspondence in their amino acid sequences into three main families with subclasses. Joyce, C.M. and Steitz, T.A. (1994) Annu. Rev. Biochem. 63, 777-822 give a summary of the motifs and conserved amino acids that were found. In prokaryotes the main distinction is made between three polymerases: polymerase I, II and III. These polymerases differ with regard to their function in the cell and with regard to

their properties. DNA polymerase I is considered to be a repair enzyme and frequently has 5'-3' as well as 3'-5' exonuclease activity. Polymerase II appears to facilitate DNA synthesis which starts from a damaged template strand and thus preserves mutations. Polymerase III is the replication enzyme of the cell, it synthesizes nucleotides at a high rate (ca. 30,000 per minute) and is considered to be very processive. Polymerase III has no 5'-3' exonuclease activity. Other properties of polymerases are due to their origin such as e.g. thermostability or processivity.

Particular properties of polymerases are desirable depending on the application. For example thermostable, high-fidelity (i.e. polymerases with proof-reading activity), processive and rapidly synthesizing polymerases are preferred for PCR. Enzymes are preferred for sequencing which do not discriminate much between dideoxy and deoxy nucleotides. In contrast the proof-reading activity of polymerases, i.e. 3'-5' exonuclease activity, is not desirable for sequencing. For some applications e.g. PCR it is desirable that the polymerase has no or little 5'-3' exonuclease activity (5' nuclease activity).

Polymerases can also differ in their ability to accept RNA as a template i.e. with regard to their reverse transcriptase (RT) activity. The RT activity may be dependent on the presence of manganese or/and magnesium ions. It is often desirable that the RT activity of the polymerase is independent of manganese ions since the reading accuracy of polymerase is decreased in the presence of manganese ions. Polymerases additionally differ in their processivity which is also a desirable property for many applications.

There is therefore a need to optimize the properties of polymerases with regard to a particular application. In the past this was often achieved by introducing mutations or by deleting functions of the polymerases.

Thus for example the 5'-3' exonuclease activity was abolished by introducing mutations (Merken, L.S. (1995) *Biochem. Biophys. Acta* 1264, 243-248) as well as by truncation (Jacobsen, H. (1974) *Eur. J. Biochem.* 45, 623-627; Barnes, W.M. (1992) *Gene* 112, 29-35). The ability of polymerases to discriminate between dideoxy and deoxynucleotides was reduced by introducing point mutations (Tabor S. and Richardson, C.C. (1995) *Proc. Natl. Acad. Sci.* 92, 6339-6343). Tabor and Richardson describe the construction of active site hybrids.

The object to provide polymerases with optimized properties was achieved by the present invention for the first time by producing polymerase chimeras by exchanging domains that are structurally and functionally independent of one another. Domains in the sense of the present invention are understood as regions which contain all essential centres or all functionally important amino acids such that the domains essentially retain their function. It is therefore also possible to exchange only parts i.e. functioning fragments of domains. Thus these domains can be referred to as functional amino acid fragments in the sense of the present invention. Furthermore the chimeras can be additionally modified by mutations or truncations. If it appears to be advantageous it is also possible to introduce mutations into the chimeras which further optimize their properties with regard to the respective application. Thus for example mutations can be introduced which reduce the ability of the polymerases

to discriminate between dideoxy and deoxy nucleotides. Alternatively desired properties such as processivity can be strengthened or introduced by introducing mutations or by truncation. The introduction of mutations or truncations can also abolish undesired properties e.g. the 5' nuclease activity.

Thus polymerase chimeras are a subject matter of the present invention which combine advantageous properties of naturally occurring polymerases with regard to a particular application. The polymerase chimeras according to the invention are composed of functional amino acid fragments of different enzymes which preferably represent domains of different enzymes. The invention surprisingly showed that the domains from the different enzymes are active in the chimera and exhibit a cooperative behaviour between the domains. The present invention also concerns general processes for the production of polymerase chimeras with optimized properties. This process according to the invention thus enables a chimera to be designed from an arbitrary combination of enzymes by exchanging domains. It is additionally preferred that the interactions at the sites of contact between the domains are further harmonized by various methods. This can for example lead to an increase in the thermostability of the chimeras. A further subject matter of the invention is a kit for the synthesis of nucleic acids which contains a chimera according to the invention.

Thermostable DNA polymerases with proof-reading function are being increasingly used in practice for PCR. The use of mixtures of *Taq* polymerase and thermostable proof-reading DNA polymerase (such as *Pfu*, *Pwo*, *Vent* polymerase) has proven to be particularly successful for

the amplification of long DNA molecules. Thus a further subject matter of the present invention was to combine the high processivity and thermostability of *Taq* polymerase with the 3'-5' exonuclease activity of another DNA polymerase in one enzyme. Hence the present invention especially concerns thermostable polymerase chimeras which have a processivity which corresponds to at least that of *Taq* polymerase and have a low error rate when incorporating nucleotides into the polymer chain during amplification due to the presence of a 3'-5' exonuclease activity (proof-reading activity). The combination of these two properties enables for example a chimera to be generated which is able to make long PCR products i.e. nucleic acid fragments which are larger than 2 kb. The chimera according to the invention is also suitable for amplifying shorter fragments.

The present invention therefore concerns in particular a polymerase chimera which is composed of functional amino acid fragments of two different polymerases wherein the first or the second polymerase has 3'-5' exonuclease activity and the polymerase chimera has 5'-3' polymerase activity as well as 3'-5' exonuclease activity. The polymerases can be naturally occurring or recombinant polymerases. The polymerase chimera according to the invention can be composed of functional amino acid fragments from two or several different polymerases. The polymerase chimera according to the invention can be composed of two or several functional amino acid fragments from the different polymerases. The amino acid sequence of the fragment can correspond to the naturally occurring sequence of the polymerase or to a sequence modified by mutations.

The amino acid fragments from which the polymerase

chimera is constructed preferably each correspond to functional polymerase domains of the first or second polymerase. A functional polymerase domain in the sense of the present invention is a region which contains all amino acids that are essential for the activity and is abbreviated as domain in the following.

The present invention concerns in particular a polymerase chimera composed of functional amino acid fragments (in short domains) from at least two different polymerases wherein the domain having polymerase activity is homologous to one polymerase and the domain having 3' exonuclease activity is homologous to another polymerase. Moreover, this chimera can additionally have 5' exonuclease activity in which case the domain having 5' exonuclease activity can be homologous to the first or to the second polymerase. However, it is also possible that the 5' exonuclease domain is partially or completely deleted or has point mutations. The polymerase chimera according to the invention can additionally have reverse transcriptase (RT) activity.

It is additionally preferred that a part of the amino acid fragments of the polymerase chimeras corresponds to a part of the amino acid sequence of Taq polymerase.

The polymerase whose domain or amino acid fragment having 3'-5' exonuclease activity has been incorporated into the chimera can for example be a Pol-I type polymerase or also a Pol-II type polymerase.

Representatives of the Pol-I type polymerase with 3'-5' exonuclease activity are for example *Escherichia coli* polymerase (Ec.1), *Salmonella* polymerase I, *Bacillus* polymerase I, *Thermosiphon* polymerase I and *Thermatoga*

neapolitana polymerase (Tne). Representatives of the Pol-II type polymerase with 3'-5' exonuclease activity are for example *Pyrococcus woessii* polymerase (Pwo), *Pyrococcus furiosus* polymerase (Pfu), *Thermococcus litoralis* polymerase (Tli), *Pyrodictum abyssi*.

Representatives of Pol-I type and Pol-II type polymerases which were mentioned as examples are described in more detail in the following:

The *Taq* DNA polymerase from *Thermus aquaticus* (*Taq* polymerase), *Escherichia coli* DNA polymerase I (*E. coli* polI) and *Thermotoga neapolitana* DNA polymerase (*Tne* polymerase) are bacterial DNA polymerases from the A family. They are DNA polymerases of the polI type since the various enzymatic activities are located in the various domains in a relatively similar manner to that found in *E. coli* polI. The *Pyrococcus woessii* DNA polymerase (*Pwo* polymerase) is, like *Thermococcus litoralis* DNA polymerase (*Vent*TM polymerase) and *Pyrococcus furiosus* DNA polymerase (*Pfu* polymerase), an archaeobacterial DNA polymerase of the B family.

Taq polymerase is described by Chien, A. et al. (1976) J. Bacteriol. 127, 1550-1557, Kaledin, A.S. et al. (1980) Biokhimiya 45, 644-651 and Lawyer, F.C. et al. (1989) J. Biol. Chem. 264, 6427-6437. It was originally isolated from the thermophilic eubacterium *Thermus aquaticus* and later cloned in *E. coli*. The enzyme has a molecular weight of 94 kDa and is active as a monomer. *Taq* polymerase is suitable for use in the polymerase chain reaction (PCR) since it has a high thermal stability (half life of 40 minutes at 95°C/5 minutes at 100°C) and a highly processive 5'-3' DNA polymerase

(polymerisation rate: 75 nucleotides per second). Apart from the polymerase activity, a 5' nuclease activity was detected by Longley et al. (1990) Nucl. Acids Res. 18, 7317-7322. The enzyme has no 3'-5' exonuclease activity so that errors occur during the incorporation of the four deoxyribonucleotide triphosphates to successively extend polynucleotide chains which interfere with the gene amplification (error rate: 2×10^{-4} errors/base, Cha, R.S. and Thilly, W.G. (1993) PCR Methods Applic. 3, 18-29). The tertiary structure of *Taq* polymerase has been known since 1995 (Kim et al., 1995, Korolev et al., 1995).

E. coli polI is described in Kornberg, A. and Baker, T.A. (1992) DNA Replication, 2nd edition, Freeman, New York, 113-165. The enzyme has a molecular weight of 103 kDa and is active as a monomer. *E. coli* polI has 5' nuclease activity and 5'-3' DNA polymerase activity. In contrast to *Taq* polymerase, it additionally has a 3'-5' exonuclease activity as a proof-reading function. *E. coli* polI and its Klenow fragment (Jacobsen, H. et al. (1974) Eur. J. Biochem. 45, 623-627) were used for PCR before the introduction of *Taq* polymerase. However, due to their low thermal stability they are less suitable since they have to be newly added to each cycle. The tertiary structure of the Klenow fragment of *E. coli* polI has been known since 1983 (Brick, P. et al., (1983) J. Mol. Biol. 166, 453-456, Ollis, D.L. et al. (1985) Nature 313, 762-766 and Beese, L.S. et al. (1993) Science 260, 352-355).

The polymerase was isolated from the thermophilic eubacterium *Thermotoga neapolitana* and later cloned in *E. coli*. The amino acid sequence of the *Tne* polymerase is similar to that of *Thermotoga maritima* DNA polymerase (UITma™ polymerase) (personal information from Dr. B.

Frey). It has a high thermal stability, 5' nuclease activity, 3'-5' exonuclease activity and 5'-3' DNA polymerase activity. A disadvantage is the low polymerisation rate compared with that of *Taq* polymerase. The UITmaTM polymerase which has a similar amino acid sequence is used for PCR if a high accuracy is required. Of the structure of *Tne* polymerase, only the amino acid sequence is known up to now (Boehringer Mannheim). However, the enzyme is homologous to *E. coli* polI so that, although the tertiary structure is unknown, homology modelling is possible.

Pfu polymerase was isolated from the hyper-thermophilic, marine archaeobacterium *Pyrococcus furiosus*. It has a high thermal stability (95 % activity after one hour at 95°C), 3'-5' exonuclease activity and 5'-3' DNA polymerase activity (Lundberg, K.S. et al. (1991) Gene 108, 1-6). The accuracy of the DNA synthesis is ca. 10 times higher than that of *Taq* polymerase. It is used for PCR if a high accuracy is required. Of the structure only the amino acid sequence is known up to now.

Pwo polymerase (PCR Applications Manual (1995), Boehringer Mannheim GmbH, Biochemica, 28-32) was originally isolated from the hyperthermophilic archaeobacterium *Pyrococcus woesei* and later cloned in *E. coli*. The enzyme has a molecular weight of about 90 kDa and is active as a monomer. *Pwo* polymerase has a higher thermal stability than *Taq* polymerase (half life > 2 hours at 100°C), a highly processive 5'-3' DNA polymerase activity and a high 3'-5' exonuclease activity which increases the accuracy of the DNA synthesis. The enzyme has no 5' nuclease activity. The polymerisation rate (30 nucleotides per second) is less than that of *Taq* polymerase. The enzyme is used for PCR if a high

accuracy is required. The accuracy of the DNA synthesis is more than 10 times higher than when using *Taq* polymerase.

Ath polymerase was isolated from the thermophilic archaeobacterium *Anaerocellum thermophilum* and later cloned in *E. coli*. *Ath* polymerase has a high thermal stability and still has at least 90 % of the original activity after an incubation of 30 min at 80°C in the absence of stabilizing detergents. The polymerase also has RT activity in the presence of magnesium ions. *Ath* polymerase is deposited at the "Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH", Mascheroder Weg 1b, D38124 Braunschweig DSM Accession No. 8995. The *Ath* polymerase has 5'-3' polymerase activity, 5'-3' exonuclease activity but no 3'-5' exonuclease activity.

Histidine tags or other purification aids can be additionally incorporated into the amino acid sequence of the polymerase chimeras to improve the purification.

There are four main methods for introducing a 3'-5' exonuclease activity of a polymerase into another polymerase for example into *Taq* polymerase which are also a subject matter of the present invention:

1. Insertion of the 3'-5' exonuclease region of another DNA polymerase by exchange of a molecular region of *Taq* polymerase

This approach is particularly suitable since the *Taq* polymerase is homologous to *E. coli* polI which is composed of domains which are functionally and structurally independent (Joyce, C.M. and Steitz, T.A.

(1987) TIBS 12, 288-292) and can serve as a model for other DNA polymerases (Joyce, C.M. (1991) Curr. Opin. Struct. Biol. 1, 123-129). Suitable DNA polymerases for the exchange are those for which a 3'-5' exonuclease activity has been demonstrated, whose DNA sequence is known and the gene coding for the 3'-5' exonuclease activity is available. For a rational protein design based on model structures it is additionally advantageous that the 3'-5' exonuclease region and the polymerase region are homologous to *E. coli* polI. The 3'-5' exonuclease region preferably fits well into the structure of *E. coli* polI and adjoins the polymerase region of *Taq* polymerase. Further advantages are an elucidated tertiary structure with available structural data and high thermal stability of the protein.

The following DNA polymerases are thus for example suitable:

a. *E. coli* polI

Apart from thermal stability, *E. coli* polI fulfils all the above-mentioned conditions. The tertiary structure of the Klenow fragment is available in the Brookhaven data bank and, like *Taq* polymerase, it belongs to the A family of DNA polymerases. The identity in the amino acid sequence is 32 %. Taking the known domain structure into consideration, the largest agreements are found in the N-terminal and in the C-terminal region of the two proteins (32 % identity in the 5' nuclease domains, 49 % identity in the polymerase domains). The shorter *Taq* polymerase has several deletions in the region of the 3'-5' exonuclease domain (14 % identity in the 3'-5' exonuclease domain and intermediate domain). Since *E. coli* polI is thermolabile and the interactions at the interface between the two domains in the chimeric

protein are no longer optimal, it is probable that the protein chimera will also have a lower thermal stability than that of *Taq* polymerase. This can be redressed by subsequent modification of amino acids at the interface.

b. Thermostable DNA polymerases

Among the thermostable DNA polymerases with 3'-5' exonuclease that are nowadays used for PCR, the *Pwo* polymerase, *Pfu* polymerase, VentTM polymerase, *Tne* polymerase and UITmaTM polymerase appear to be suitable for combination with the *Taq* DNA polymerase. The genes of the *Pwo* polymerase and the *Tne* polymerase are accessible (via the Boehringer Mannheim Company). The *Pfu* polymerase can be obtained from Stratagene Inc. The *Tne* polymerase is well suited for a rational protein design due to its homology to *Taq* polymerase and *E. coli* polI. When using the *Pfu* polymerase designs are only possible based on amino acid sequence alignments taking into consideration the known conserved amino acids and motifs that are essential for the function.

2. Modification of the *Taq* DNA polymerase in the intermediate domain

In order to insert a 3'-5' exonuclease activity it is necessary to insert all amino acids that are essential for the activity into the structure. According to the present state of knowledge this applies in particular to the three motifs Exo I, Exo II and Exo III. The essential motifs must additionally be linked in a suitable manner in order to be placed in the spatial position necessary for catalysis.

It is also possible to modify the *Taq* DNA polymerase in the polymerase region. A de novo design of polymerases is also in principle conceivable.

The chimeras according to the invention can be additionally optimized by:

1. Removing the 5' nuclease domain (possible also proteolytically) or subsequently inactivating the 5' nuclease activity (described in Merkens, L.S. (1995) *Biochem. Biophys. Acta* 1264, 243-248)
2. Modification by point mutations or fragment exchange
3. Optimization of the structures at the interface of the chimeras
4. Optimization by random mutagenesis and/or random recombination with other polymerase genes (molecular evolution).

Examples of polymerase chimeras according to the invention are the following:

- *Taq* DNA polymerase (M1-V307)*E.coli* DNA polymerase (D355-D501) *Taq* DNA polymerase (A406-E832)
- *Taq* DNA polymerase (M1-P291)*E.coli* DNA polymerase (Y327-K511) *Taq* DNA polymerase (L416-E832)
- *Taq* DNA polymerase (M1-P291)*E.coli* DNA polymerase (Y327-H519) *Taq* DNA polymerase (E424-E832): point mutation A643G; Ile455Val SEQ ID NO.:1
- *Taq* DNA polymerase (M1-P291)*E.coli* DNA polymerase (Y327-V536) *Taq* DNA polymerase (L441-E832)
- *Taq* DNA polymerase (M1-P291)*E.coli* DNA polymerase (Y327-G544) *Taq* DNA polymerase (V449-E832); SEQ ID NO.:2
- *Taq* DNA polymerase (M1-P302)*E.coli* DNA polymerase (K348-S365) *Taq* DNA polymerase (A319-E347) *E.coli* DNA

- poly(N450-T505) *Taq* DNA polymerase (E410-E4832);
- *Taq* DNA polymerase (M1-V307)*Tne* DNA polymerase (D323-D468) *Taq* DNA polymerase (A406-E832)
 - *Taq* DNA polymerase (M1-P291)*Tne* DNA polymerase (P295-I478) *Taq* DNA polymerase (L416-E832)
 - *Taq* DNA polymerase (M1-P291)*Tne* DNA polymerase (P295-E485) *Taq* DNA polymerase (E424-E832); silent mutation A1449C SEQ ID NO.:3
 - *Taq* DNA polymerase (M1-P291)*Tne* DNA polymerase (P295-V502) *Taq* DNA polymerase (L441-E832)
 - *Taq* DNA polymerase (M1-P291)*Tne* DNA polymerase (P295-G510) *Taq* DNA polymerase (V449-E832); silent mutation C1767T SEQ ID NO.:4
 - *Taq* DNA polymerase (M1-P302)*Tne* DNA polymerase (E316-D333) *Taq* DNA polymerase (A319-E347) *Tne* DNA polymerase (I381-M394) *Taq* DNA polymerase (R362-L380) *Tne* DNA polymerase (E415-T472)*Taq* DNA polymerase (E410-E832); G308D/V310E/L352N/L356D/E401Y/R305D
 - *Taq* DNA polymerase (1-291)*Pfu* DNA polymerase (V100-R346) *Taq* DNA polymerase (E424-E832)
 - *Taq* DNA polymerase (1-291)*Pfu* DNA polymerase (H103-S334) *Taq* DNA polymerase (E424-E832); SEQ ID NO.:5
 - *Taq* DNA polymerase (1-291)*Pfu* DNA polymerase (V100-F389) *Taq* DNA polymerase (E424-E832)
 - *Taq* DNA polymerase (1-291)*Pfu* DNA polymerase (V100-F389) *Taq* DNA polymerase (V449-E832); SEQ ID NO.:6
 - *Taq* DNA polymerase (1-291)*Pfu* DNA polymerase (M1-F389) *Taq* DNA polymerase (V449-E832)

Of the above-mentioned polymerase chimeras the following were examined in more detail:

- *Taq* DNA polymerase (M1-P291)*E.coli* DNA polymerase (Y327-H519) *Taq* DNA polymerase (E424-E832): point mutation A643G; Ile455Val (*Taq* Ec1) SEQ ID NO.:1

- *Taq* DNA polymerase (M1-P291)*E.coli* DNA polymerase (Y327-G544) *Taq* DNA polymerase (V449-E832), (*Taq* Ec2) SEQ ID NO.:2
- *Taq* DNA polymerase (M1-P291)*Tne* DNA polymerase (P295-E485) *Taq* DNA polymerase (E424-E832); silent mutation A1449C (*Taq* *Tne*1) SEQ ID NO.:3
- *Taq* DNA polymerase (M1-P291)*Tne* DNA polymerase (P295-G510) *Taq* DNA polymerase (V449-E832); silent mutation C1767T (*Taq* *Tne*2) SEQ ID NO.:4
- *Taq* DNA polymerase (1-291)*Pfu* DNA polymerase (V100-R346) *Taq* DNA polymerase (E424-E832), (*Taq* *Pfu*1) SEQ ID NO.:5
- *Taq* DNA polymerase (1-291)*Pfu* DNA polymerase (V100-F389) *Taq* DNA polymerase (V449-E832), (*Taq* *Pfu*2) SEQ ID NO.:6

In order to select suitable DNA polymerases, multiple amino acid sequence alignments of available sequences of DNA polymerases and DNA binding proteins are established for example with the program GCG (Devereux et al., 1984, Nucl. Acids Res. 12, 387-395). In order to find a good alignment it is necessary to take into consideration the secondary structure predictions, known structure-based sequence alignments, known motifs and functionally essential amino acids as well as phylogenetic aspects. If the proteins are composed of functionally and structurally independent domains it is appropriate to firstly establish the amino acid sequence alignments with respect to the individual domains and only afterwards to combine them into a complete sequence alignment.

If homologous sequences are found whose tertiary structure is known, then it is possible to derive a 3D model structure from the homologous protein. The program BRAGI (Reichelt and Schomburg, 1988, J. Mol. Graph. 6, 161-165) can be used to make the model. The program

AMBER (Weiner et al., 1984, J. Am. Chem. Soc. 106, 765-784) can be used for energy minimization of the structures of individual molecule regions and whole molecules and the program Procheck can be used to check the quality of the model. If only the C α coordinates of the structure of the initial protein are available, the structure can for example be reconstructed using the program O (Jones et al., 1991, Acta Cryst. A47, 110-119). It is also possible to obtain C α coordinates that are not available in the protein data bank but have been already published as a stereo picture by scanning the stereo picture and picking out the coordinates (for example using the program Magick) and calculating the z-coordinates (for example using the program stereo). Variants can be designed based on amino acid sequence alignments, based on 3D models or based on experimentally determined 3D structures.

In addition chimera variants were produced in which the domain with polymerase activity has reverse transcriptase activity. Examples of suitable polymerases are e.g. the polymerase from *Anaerocellum thermophilum* Ath or *Thermus thermophilum* Tth. The 3'-5' exonuclease activity is inserted by a domain which is derived from another polymerase e.g. the Tne polymerase or the Pfu or Pwo polymerase. This chimera can additionally have 5'-3' exonuclease activity in which case the domain with 5' exonuclease activity can be derived from the first as well as from the second polymerase.

The recombinant hybrid polymerases HYB and HYBd5, like the DNA polymerase from *Anaerocellum thermophilum*, have a relatively strong reverse transcriptase activity in the presence of magnesium ions as well as in the presence of manganese ions. As shown in Figure 22 the

ratio of polymerase activity to reverse transcriptase activity is more favourable than with the Tth polymerase which is the most common and well-known enzyme of this type. This finding applies to the magnesium-dependent as well as to the manganese-dependent reverse transcriptase activity. It can be concluded from this that the polymerase domain which is derived from the *Anaerocellum* polymerase also exhibits full activity in the hybrid enzyme. The variant HYBd5 additionally has 3'-5' exonuclease activity as shown in Figure 21. This is inhibited by the presence of deoxynucleoside triphosphates as expected for the typical "proof-reading activity". The exonuclease domain which is derived from the DNA polymerase from *Thermotoga neapolitana* is thus also active in the hybrid molecule. The ability to inhibit the exonuclease activity also demonstrates that both domains of the hybrid polymerase molecule interact and thus the hybrid polymerase is functionally very similar to the natural enzyme.

The production of domain exchange variants by genetic engineering can be achieved by PCR mutagenesis according to the SOE method (Horton et al. (1989) *Gene* 77, 61-68) or by the modified method (cf. scheme in the examples) with the aid of chemically synthesized oligodeoxynucleotides. The respective DNA fragments are separated on an agarose gel, isolated and ligated into the starting vector. pUC derivatives with suitable promoters such as pTE, pTaq, pPL, Bluescript can be used as starting vectors for *E. coli*. The plasmid DNA is transformed into an *E. coli* strain, for example XL1-blue, some clones are picked out and their plasmid DNA is isolated. It is also possible to use other strains such as Nova Blue, BL21 (DE), MC1000 etc. Of course it is also possible to clone into other organisms such as

into yeast, plant and mammalian cells. A preselection of clones whose plasmid DNA is sequenced in the modified region is made by restriction analysis.

The gene expression in the target proteins can be induced by IPTG in many plasmids such as P_btaq. When producing many different variants it is appropriate to establish a universal purification procedure. Affinity chromatography on Ni-NTA (nickel-nitrilotriacetic acid) agarose is well suited for this which can be used after attaching a His tag to the protein, for example by PCR. The protein concentrations can be determined with the protein assay ESL (Boehringer Mannheim) and contaminating side activities of the preparations can be determined as described for the commercially available Taq polymerase (Boehringer Mannheim). Polymerase, exonuclease activity and thermostability tests are carried out to further characterize the variants and the respective temperature optimum is determined. The polymerase activities of the chimeras can be determined in non-radioactive test systems for example by determining the incorporation rate of Dig-dUTP into DNase activated calf thymus DNA, or in radioactive test systems by for example determining the incorporation rate of α -[³²P]dCTP into M13 mp9 ssDNA. In order to determine the temperature optima of the polymerase activity of the chimeras, the polymerase reaction is carried out at different temperatures and the specific activities are calculated. The residual activities (i.e. the percentage of the initial activity without heat treatment) after heat treatment are measured in order to determine the thermal stabilities. The 3'-5' exonuclease activity can be demonstrated by incorporation of a 5'-Dig-labelled primer which anneals to a DNA template strand starting at its 3' end. The correction of 3'

mismatched primers and their extension (proof reading) can be shown by the extension of mismatched 5'-Dig-labelled primers which anneal to a template strand in the recognition sequence of a restriction enzyme (e.g. EcoRI). A cleavage with the restriction enzyme is only possible when the mismatch is corrected by the enzyme. The processivity can be examined by using variants in the PCR. If the enzyme is not sufficiently thermostable for use in PCR, a PCR can be carried out at the temperature optimum as the extension temperature with successive addition of enzyme. The exonuclease activity of the chimeras can be determined in a radioactive test system. For this a certain amount of the chimeric polymerases (usually 2.5 U) is incubated for 4 hours at various temperatures with labelled DNA (5 μ g [3 H] DNA in the respective test buffers). dNTPs were optionally added at various concentrations (0 - 0.2 mM). After terminating the reaction the release of radioactively labelled nucleotides is determined.

A further subject matter of the present invention is the DNA sequence of the polymerase chimeras described above. In particular the DNA sequences SEQ ID NO.: 1-6 are a subject matter of the present invention. The present invention additionally concerns the amino acid sequences of the polymerase chimera described above. In particular the amino acid sequences SEQ ID NO.: 7-12 are a subject matter of the present invention. Moreover the DNA sequence SEQ ID NO.:17 is a subject matter of the invention.

Vectors which contain the above-mentioned DNA sequences are a further subject matter of the present invention. pBTaq (plasmid Pbtac4_oligo 67 (Villbrandt (1995), dissertation, TU Braunschweig)) is a preferred vector.

The *E. coli* strains, in particular the strain *Escherichia coli* XL1-blue which contain the vector which carries the polymerase chimera gene are a further subject matter of the invention. The following strains were deposited at the DSM, "Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH", Mascheroder Weg 1b, D-38124 Braunschweig:

- *E.coli* XL1 Blue x pBTaqEc1: **TaqEc1** DSM No. 12053
- *E.coli* XL1 Blue x pBTaqTne1:**TaqTne1** DSM No. 12050
- *E.coli* XL1 Blue x pBTaqTne2:**TaqTne2** DSM No. 12051
- *E.coli* XL1 Blue x pBTaqPfu1:**TaqPfu1** DSM No. 12052

The polymerase chimeras according to the invention are particularly suitable for amplifying DNA fragments e.g. for the polymerase chain reaction. A further application is for example to sequence DNA fragments.

A preferred vector for the Ath-Tne chimera is the following:

E.coli BL 21 (DE3) plySS x pETHYBR : HYBR

E.coli BL 21 (DE3) plySS x pETHYBR d5: HYBR d5

The *E. coli* strains which contain the vector which carries the polymerase chimera gene are a further subject matter of the invention. The following strains were deposited at the DSM, "Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH", Mascheroder Weg 1b, D-38129 Braunschweig: HYBR (DSM No. 12720); HYBR d5 (DSM No. 12719).

The production of the above-mentioned Ath-Tne chimeras is described for example in examples 8-11. The chimeras according to the invention which have RT activity are

particularly suitable for the reverse transcription of RNA.

A further subject matter of the present invention is a kit for amplifying DNA fragments which contains at least one of the polymerase chimeras according to the invention.

Short description of the Figures

Figure 1:

DNA sequence of the *Taq* DNA polymerase (M1-P291) *E. coli* DNA polymerase (Y327-H519) *Taq* DNA polymerase (E424-E832): point mutation A643G; Ile455Val SEQ ID NO.:1; and the corresponding amino acid sequence SEQ ID NO.:7.

Figure 2:

DNA sequence of the *Taq* DNA polymerase (M1-P291) *E. coli* DNA polymerase (Y327-G544) *Taq* DNA polymerase (V449-E832); SEQ ID NO.:2; and the corresponding amino acid sequence SEQ ID NO.:8.

Figure 3:

DNA sequence of the *Taq* DNA polymerase (M1-P291) *Tne* DNA polymerase (P295-E485) *Taq* DNA polymerase (E424-E832); silent mutation A1449C SEQ ID NO.:3; and the corresponding amino acid sequence SEQ ID NO.: 9.

Figure 4:

DNA sequence of the *Taq* DNA polymerase (M1-P291) *Tne* DNA polymerase (P295-G510) *Taq* DNA polymerase (V449-E832); silent mutation C1767T SEQ ID NO.:4; and the corresponding amino acid sequence SEQ ID NO.:10.

Figure 5:

DNA sequence of the *Taq* DNA polymerase (1-291) *Pfu* DNA polymerase (H103-S334) *Taq* DNA polymerase (E424-E832); SEQ ID NO.:5; and the corresponding amino acid sequence SEQ ID NO.:11.

Figure 6:

DNA sequence of the *Taq* DNA polymerase (1-291) *Pfu* DNA polymerase (V100-F389) *Taq* DNA polymerase (-V449-E832); SEQ ID NO.:6; and the corresponding amino acid sequence SEQ ID NO.:12.

Figure 7:

Purification of the domain exchange variant *Taq*Ec1 on Ni-NTA agarose. Analysis on an 8 % polyacrylamide gel stained with Coomassie blue.

Lanes: 1,8 protein molecular weight marker Broad Range
 (200 kDa, 116.25 kDa, 97.4 kDa, 66.2 kDa,
 45 kDa, 31 kDa)
lane 2 soluble proteins
lane 3 column flow-through
lane 4 wash fraction buffer B
lane 5 wash fraction buffer A
lanes 6,7 eluate fraction buffer C
protein yield (OD₂₈₀) about 7 mg

Figure 8:

Determination of protein purity: SDS-PAGE, Phast system (10-15 %): silver staining MW: protein molecular weight markers; NHis-TaqPol: *Taq* DNA polymerase with N-terminal His tag; *Taq*Ec1, *Taq*Tne1, *Taq*Tne2: domain exchange variants.

Figure 9:

Specific activities of the domain exchange variants at various temperatures.

Figure 10:

Testing the domain exchange variants in the PCR with successive addition of enzyme, extension at 72°C.

lambda DNA (left): size of the target sequence = 500 bp
plasmid pa (right): size of the target sequence = 250 bp
lane 1: Taq DNA polymerase (BM Co.), 100 ng, 5 units
lane 2: domain exchange variant TaqEc1, 500 ng,
1.25 units/cycle
lane 3: domain exchange variant TaqTne1, 50 ng,
3.6 units/cycle
lane 4: domain exchange variant TaqTne2, 50 ng,
3.5 units/cycle
III: DNA length standard III (BM Co.)
VI: DNA length standard VI (BM Co.).

Result: When the domain exchange variant TaqTne2 was used, PCR products of the correct size were formed.

Figure 11:

Testing the domain exchange variants in the PCR with successive addition of enzyme, extension at 55°C.

lambda DNA (left): size of the target sequence = 500 bp
plasmid pa (right): size of the target sequence = 250 bp
lane 1: domain exchange variant TaqEc1, 500 ng,
6 units/cycle
lane 2: domain exchange variant TaqTne1, 50 ng,
7.5 units/cycle
III: DNA length standard III (BM Co.)

VI: DNA length standard VI (BM Co.).

Result: When the domain exchange variant TaqEc1 was used, PCR products of the correct size were formed.

Figure 12:

3'-5' exonuclease test-variant TaqEc1, incubation at 72°C, primer P1.

Figure 13:

3'-5' exonuclease test-variant TaqEc1, incubation at 50°C, primer P1 (left), primer P2 (right).

Figure 14:

Correction of 3' mismatched primers and their extension - variant TaqEc1 (3' mismatch primer correction assay)
(-): without restriction enzyme digestion
(+): restriction enzyme digestion with EcoRI.

Figure 15:

Schematic representation
Degradation of primers at the 3' end (3'-5' exonuclease assay) and correction of 3' mismatched primers and their extension (3' mismatch primer correction assay).

Figure 16:

Schematic representation: simplified flow chart, degradation of primers at the 3' end and correction of 3' mismatched primers and extension.

Figure 17:

CLUSTAL W (1.5) multiple sequence alignment of the Ath, Tne, PolI polymerase genes as well as of the predicted

gene of the polymerase chimera. The part of the chimera sequence which is derived from Tne is underlined.

Figure 18:

- A. Structure of the primers which were used for the PCR amplification of the Tne-Exo and the Ath polymerase domains.
- B. Part of the amino acid sequence alignment of two polymerases which exhibited the selected crossing point.
- C. Nucleotide sequence and position of the primers which were designed for the construction of the hybrid polymerase gene. The sequences of the primers which are not complementary to the target sequence are shown in small letters. Complementary "overlapping" sequences in the TNELOW and ATHUP primers are double underlined.

Figure 19:

- A. Part of the alignment of the Ath and Tne amino acid sequences which show the homologous region that was used to splice together the domains of the two polymerases.
- B. Nucleotide and amino acid sequence of the two polymerases in the splicing region. The figure shows the single BamHI cleavage site in the Tne DNA sequence and the sequence of the two oligos that were constructed in order to introduce the BamHI cleavage site into the Ath polymerase.

Figure 20:

Construction of the gene of the polymerase chimera (cf. also example 8).

Figure 21:

3'-5' exonuclease activity of the recombinant DNA polymerase.

- 1-DNA of the lambda phage hydrolyzed by HindIII
- 2-DNA of the lambda phage hydrolyzed by HindIII, and dNTP, and recombinant DNA polymerase
- 3-DNA of the lambda phage hydrolyzed by HindIII, without dNTP, with recombinant DNA polymerase
- 4-DNA of the lambda phage hydrolyzed by HindIII.

Figure 22:

Reverse transcriptase activity of the recombinant polymerases HYB and HYBd5. The DNA polymerase activity of a 2 μ l extract from E. coli BL21 (DE3) plysS x pETHYBr and E. coli BL21 (DE3) plysS x pETHYBRd5 was determined with a precision of 0.05 units. These amounts were used to determine the reverse transcriptase activity of the hybrid polymerases and the effect of 1 mM manganese or 4 mM magnesium ions. The controls were Tth (0.25 units) as a manganese-dependent reverse transcriptase and C. therm. polymerase (Roche Molecular Biochemicals) as a magnesium-dependent reverse transcriptase.

Example 1: Construction and cloning

Establishing a universal purification procedure

Affinity chromatography on Ni-NTA (nickel-nitrilotriacetic acid) agarose was used to standardize the

purification protocol for the domain exchange variants. Before producing the protein variants it was necessary to attach or insert a His tag to or into the *Taq* DNA polymerase. Two different His tag variants in the plasmid Pbtaq4_oligo67 (Boehringer Mannheim) were designed and produced. The variant NHis-*Taq*Pol contains an N-terminal His tag, an enterokinase cleavage site to optionally cleave the His tag and an epitope for the detection of His tag proteins with antibodies (Quiagen). It was produced by PCR from the *Eco*RI site up to the *Pst*I site. In the N-terminal protein sequencing the twenty N-terminal amino acids of the variant NHis-*Taq*Pol were confirmed as correct.

Sequence: NHis-*Taq*Pol

EcoRI codon from TaqPol

5'G AA TTC ATG AGG GGC TCG CAT CAC CAT CAC CAT CAC GCT GCT GAC GAT GAC GAT AAA ATG AGG GGC 3'

Met Arg Gly Ser His His His His His His Ala Ala Asp Asp Asp Asp Lys Met Arg Gly

MRGS'His epitope (Met-Arg-Gly-Ser-(His)₆] enterokinase [(Asp)₄-Lys-X]

SEQ ID No.:13: 5' G AA TTC ATG AGG GGC TCG CAT CAC CAT
CAC CAT CAC GCT GCT GAC GAT GAC GAT AAA ATG AGG GGC 3'

SEQ ID No.:14: Met Arg Gly Ser His His His His His His
Ala Ala Asp Asp Asp Asp Lys Met Arg Gly

The variant 5DHis-*Taq*Pol contains a His tag in a flexible loop of the 5' nuclease domain between glycine 79 and glycine 80 of the *Taq* DNA polymerase and was produced by PCR mutagenesis from the *Eco*RI site up to the *Pst*I site.

Sequence: 5DHis-*Taq*Pol

SEQ ID No.: 15

SEQ ID No.: 16

5' GAG GCC TAC GGG CAT CAC CAT CAC CAT CAC GGG TAC AAG GCG 3'
GluAlaTyrGlyHisHisHisHisHisHisGlyTyrLysAla

The correctness of the plasmid DNA in each modified region of the two new genes was confirmed by DNA sequencing. Both modified genes were expressed under the same conditions and at the same rate as the initial protein without a His tag, they could be readily purified by Ni-NTA agarose and behaved like *Taq* polymerase without a His tag in the standard PCR. The N-terminal His tag was used to purify the domain exchange variants.

Amino acid sequence alignments

The following amino acid sequence alignments were set up in order to design the domain exchange variants:

1. *Tne*, *E. coli* I and *Taq* DNA polymerase
2. *Pfu*, *E. coli* I and *Taq* DNA polymerase
3. Multiple amino acid sequence alignments of DNA polymerases

The alignments were established with the program GCG with reference to individual molecule regions (domains) and assembled to form the complete sequence alignment taking into consideration the known secondary structures, motifs and essential amino acids and using the structure-based sequence alignment of the sequences of the 3'-5' exonuclease domain of the Klenow fragment with the corresponding domain of *Taq* DNA polymerase (figure 2d in Kim et al. (1995) Nature 376, 612-616).

In order to select the initial structure of the Klenow fragment for the homology modelling, the structures of

E. coli DNA polymerase I that were available at that time were compared using the program Bragi and an RMS fit:

Klenow fragment-dCMP complex (PDB code: 1dpi), 2.8 Å (1987), Klenow fragment-dCTP complex (PDB-code: 1kfd) 3.9 Å (1993) and Klenow fragment, D355A - DNA complex (PDB-code: 1 kln) 3.2 Å (1994).

The structure Klenow fragment (PDB-code: 1 kln) was selected. Two loops were incorporated into the two regions in which there were no coordinates (Bragi program) and energy-minimized (Amber program). The quality of the protein structure was checked (Procheck program).

Construction of 3D models

A 3D model of the molecular region of the *Taq* DNA polymerase which comprises amino acids 292-832 was constructed using the Bragi program in homology to the structure of the Klenow fragment (PDB-code: 1 kln). The modelling comprised amino acid substitutions, introduction of insertions and deletions, energy-minimization of the new loop regions and energy-minimization of the entire molecule (Amber program).

The structure of *Taq* DNA polymerase was already published at the time of the modelling work but was not available in the protein data bank. In order to set up a model of the intermediate domain of the *Taq* DNA polymerase which corresponds to the 3'-5' exonuclease domain of the Klenow fragment (amino acids 292-423), a stereo picture (Figure 2c in Kim et al. (1995) Nature **376**, 612-616) was scanned, the C α coordinates were picked out on the screen (x and y coordinates for the

left and right picture) (Magick program, (John Cristy, E.I. du Pont De Nemours and Company Incorporated)), the z coordinates were calculated (Stereo program, (Collaborative Computational Project, Number 4 (1994) Acta Cryst. D50, 760-763)), the protein main chain was reconstructed with generation of a poly-alanine (program O), amino acid substitutions were carried out (Bragi program) and an energy-minimization of the entire molecule was carried out (Amber program). The model of the amino acid residues 292-423 (see above) was added to the model of the polymerase domain (amino acids 424-832) (see above) while allowing for the structural alignments of the *Taq* DNA polymerase with the Klenow fragment (Figure 2b and 2c in Kim et al. (1995) Nature 376, 612-616). The entire model structure was energy-minimized (Amber program) and the quality of the model structure was checked (Procheck program, (Laskowski, R., A., et al. (1993) J. Appl. Cryst. 26, 283-291)).

A 3D model of the *Tne* DNA polymerase (residues 297-893) was set up in homology to the structure of the Klenow fragment (PDB-code: 1kln). The modelling included amino acid substitutions, introduction of insertions and deletions (Bragi program), energy-minimization of the new loop regions, energy-minimization of the entire molecule (Amber program) and checking the quality of the model structure (Procheck program).

20 Protein variants were designed.

They were based on the 3D structure models when using *E. coli* polI and *Tne* polymerase, and based on the amino acid alignments when using the *Pfu* polymerase.

Production of the domain exchange variants by genetic engineering

The N-terminal His tag was inserted by PCR and the domain exchange variants were produced by a modified SOE method (Horton et al. (1989) Gene 77, 61-68), shown in the scheme with the aid of chemically synthesized oligodeoxynucleotides. The respective DNA fragments were separated on an agarose gel, isolated using the QIAquick gel extraction kit (Qiagen company) according to the protocol supplied and used in PCR reactions I to IV in the subsequent PCR reaction or in the case of the PCR reaction V they were recleaved with the two restriction enzymes whose recognition sequence was located in the flanking primers (EcoRI and Pst I). The ligation of DNA fragments and the production and transformation of competent XL1 Blue *E. coli* cells by electroporation was carried out as described by Villbrandt (1995, Dissertation, TU Braunschweig). Several clones were picked out and their plasmid DNA was isolated according to the protocol supplied using the QIAprep Spin Plasmid Kit (Qiagen company). Microbiological working techniques and the formulations for preparing liquid or plate media as well as the establishment of glycerin cultures was carried out as described in the handbook by Sambrook et al. (1989, Molecular cloning - a laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The domain exchange variants were expressed at the same rate as the initial protein.

Example 2: Purification (for one chimera)

Purification of the domain exchange variants

All domain exchange variants were isolated by the same protocol from *Escherichia coli* XL1-Blue. The fermentation was carried out for 16 hours at 37°C on a

one litre scale in LB medium/100 mg/ml ampicillin/12.5 mg/ml tetracycline/1 mM IPTG. The cells were centrifuged, taken up in 20 ml lysis buffer (50 mM Tris-HCl, pH 8.5, 10 mM 2-mercaptoethanol, 1 mM PMSF), frozen at -70°C for at least 16 hours and treated for 10 minutes with ultrasound. The cell debris was centrifuged and the sterile-filtered supernatant was applied to an Ni-NTA (nickel-nitriloacetic acid) agarose column (Qiagen) with a column volume of 3.5 ml ($r=0.65$ cm, $h=2.7$ cm). It was washed with 40 ml buffer A (20 mM Tris-HCl, pH 8.5, 100 mM KCl, 20 mM imidazole, 10 mM 2-mercaptoethanol, 10 % (v/v) glycerol), subsequently with 10 ml buffer B (20 mM Tris-HCl, pH 8.5, 1 M KCl, 20 mM imidazole, 10 mM 2-mercaptoethanol, 10 % (v/v) glycerol) and again with 10 ml buffer A. It was eluted with 15 ml buffer C (20 mM Tris-HCl, pH 8.5, 100 mM KCl, 100 mM imidazole, 10 mM 2-mercaptoethanol, 10 % (v/v) glycerol). The flow rate was 0.5 ml/minute and the fraction size was 10 ml with the wash fractions and 1 ml for the elution fractions. The combined fractions were dialysed against storage buffer (20 mM Tris-HCl pH 8.0, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 % Tween 20, 50 % glycerol) and 200 μ g/ml gelatin and Nonidet P40 at a final concentration of 0.5 % were added. The protein solutions were stored at -20°C.

The analysis of the purification of the domain exchange variant TaqEc1 on Ni-NTA agarose is shown in Fig. 7.

Determination of the protein concentration

The protein concentrations were determined by measuring the OD₂₈₀ and with the protein assay ESL (Boehringer Mannheim). Figure 8 shows the determination of the protein purity: SDS-PAGE, Phast system (10-15 %): silver staining.

**Example 3: Temperature optimum of the polymerase
activity of the chimeras**

The polymerase activities of the chimeras were determined in a non-radioactive test system. A radioactive test system was used to adjust the values. The incorporation rate of Dig-dUTP into DN'ase-activated calf thymus DNA was determined in the non-radioactive test system. A 50 μ l test mix contained 5 μ l buffer mix (500 mM Tris-HCl, 150 mM $(\text{NH}_4)_2\text{SO}_4$, 100 mM KCl, 70 mM MgCl_2 , 100 mM 2-mercaptoethanol, pH 8.5), 100 μ M each of dATP, dCTP, dGTP, dTTP, 36 nM Dig-dUTP (Boehringer Mannheim), 12 μ g calf thymus DNA (DN'ase-activated), 10 μ g bovine serum albumin and 2 μ l chimeric enzyme or 0.02 units *Taq* polymerase (Boehringer Mannheim) as a reference in dilution buffer (20 mM Tris-HCl, pH 8.0, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 200 μ g/ml gelatin, 0.5 % Tween 20, 0.5 % Nonidet P40, 50 % glycerol). The reaction mixtures were incubated for 30 minutes at various temperatures. The reactions were stopped on ice. 5 μ l of each reaction mixture was pipetted into white membrane-coated microtitre plates (Pall BioSupport, SM045BWP) and baked for 10 minutes at 70°C. The membrane of the microtitre plate was treated as follows using the accompanying suction trough (Pall Bio Support): apply 100 μ l buffer 1 (1 % blocking reagent (Boehringer Mannheim) in 0.1 M maleic acid, 0.15 M NaCl, pH 7.5), incubate for 2 minutes, suck through, repeat once; apply 100 μ l buffer 2 (1:10000 diluted anti-Dig-AP-Fab fragment antibodies (Boehringer Mannheim) in buffer 1), incubate for 2 minutes, suck through, repeat once; apply 200 μ l buffer 3 (buffer 1 containing 0.3 % Tween 20) under vacuum, repeat once; apply 200 μ l buffer 4 (0.1 M Tris-HCl, 0.1 M NaCl, 50 mM MgCl_2 , pH 9.5) under vacuum; apply 50 μ l buffer 5 (1:100 diluted CSPD (Boehringer

Mannheim) in buffer 4), incubate for 5 minutes, suck through. The samples were measured in a luminometer (Microluminar LB 96P, Berthold or Wallac Micro Beta Trilux).

In the radioactive test system the incorporation rate of α -[^{32}P]dCTP into 1 μg M13mp9 ss-DNA was determined. A 50 μl test mix contained 5 μl buffer mix (670 mM Tris-HCl, 50 mM MgCl_2 , 100 mM 2-mercaptoethanol, 2 % Tesit, 2 mg/ml gelatin, pH 8.8), 10 μM each of dATP, dGTP, dTTP, 5 μM CTP, 0.1 μCi [α - ^{32}P]dCTP, 1 μg M13mp9ss DNA annealed with 0.3 μg M13 primer and 1 μl chimeric enzyme or 0.01 units *Tag* polymerase (Boehringer Mannheim) as a reference in dilution buffer (20 mM Tris-HCl, pH 8.0, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 200 $\mu\text{g}/\text{ml}$ gelatin, 0.5 % Tween 20, 0.5 % Nonidet P40, 50 % glycerol). In order to prepare the DNA primer mixture, 277.2 μg M13mp9ssDNA (Boehringer Mannheim) and 156 μg M13 sequencing primer (17mer) were heated for 30 minutes to 55°C and cooled for 30 minutes to room temperature. The reaction mixtures were incubated for 30 minutes at 65°C. The reactions were stopped on ice. 25 μl of each of the reaction solutions was removed and pipetted into 250 μl 10 % trichloroacetic acid (TCA)/0.01 M sodium pyrophosphate (PPi), mixed and incubated for 30 minutes on ice. The samples were aspirated over pre-soaked GFC filters (Whatman), the reaction vessels were washed out with 5 % TCA/PPi and the filters were washed at least three times with the same solution. After drying, the filters were measured in a β -counter in 5 ml scintillation liquid. The enzyme samples were diluted in enzyme dilution buffer. 1 μl aliquots of the dilutions were used. Duplicate or triplicate determinations were carried out. The *Tag* DNA polymerase from the Boehringer Mannheim Company was used as a reference.

One unit is defined as the amount of enzyme that is necessary to incorporate 10 nM deoxyribonucleotide triphosphate into acid-precipitable DNA at 65°C in 30 minutes. In order to determine the standard values, 2 μ l aliquots of the total mixture were pipetted onto a dry filter and dried. The blank value was determined by also incubating samples without enzyme and washing them identically.

The temperature optima were determined using the non-radioactive DNA polymerase test at various temperatures.

Specific activities at various temperatures

| Enzyme | Temperature [°C] | | | | | |
|-------------|------------------|---------|---------|---------|---------|---------|
| | 25 | 37 | 50 | 60 | 72 | 80 |
| TaqPol (BM) | 0.0 | 0.0 | 5764.4 | 8489.1 | 50000.0 | 57986.1 |
| NHIs-TaqPol | 0.0 | 0.0 | 5616.1 | 12165.2 | 60843.7 | 74784.4 |
| TaqEc1 | 704.9 | 10353.4 | 50066.5 | 41034.4 | 2677.5 | 1016.2 |
| TaqTne1 | 0.0 | 2559.4 | 15967.0 | 18900.4 | 1100.0 | 0.0 |
| TaqTne2 | 747.2 | 5180.2 | 23549.6 | 30627.3 | 64139.1 | 28727.4 |

Example 4: Temperature stability of the polymerase activity of the chimeras

The thermal stability was determined by heating the reaction mixtures to 80°C and 95°C for one, three or six minutes and subsequently determining the residual activities using the non-radioactive DNA polymerase test (see Fig. 9).

Table: residual activities (percent of the initial activity without heat treatment) at 72°C of the Taq DNA polymerase (TaqPol), the Taq DNA polymerase with a His

tag (NHis-TaqPol) and the three domain exchange variants (TaqEc1, TaqTne1, TaqTne2) after heat treatment (incorporation of Dig-dUTP into DNase-activated calf thymus DNA).

| Enzyme | 1min 80°C | 3 min 80°C | 6 min 80°C | 1 min 95°C | 3 min 95°C | 6 min 95°C |
|-------------|--------------|---------------|---------------|---------------|---------------|---------------|
| TaqPol | 100 | 100 | 100 | 100 | 100 | 100 |
| NHis-TaqPol | 100 | 100 | 100 | 100 | 100 | 100 |
| TaqEc1 | 0 | 0 | 0 | 0 | 0 | 0 |
| TaqTne1 | 16 | 0 | 0 | 0 | 0 | 0 |
| TaqTne2 | 100 | 100 | 100 | 92 | 0 | 0 |

Example 5: PCR with successive addition of enzyme

The polymerase chimeras were tested in a PCR with successive addition of enzyme. The extension was carried out at 72°C (Fig. 10) and at 55°C (Fig. 11). Each of the reactions mixtures with a reaction volume of 100 μ l contained 1 ng lambda DNA or pa-plasmid DNA (BM Co.), 1 μ M of each primer (25-mer), 200 μ M of each of the dNTPs and standard PCR buffer containing MgCl₂ (Boehringer Mannheim). The reaction conditions were:

For extension at 72°C: 1 minute 94°C / 30 seconds 50°C / 1 minute 72°C // 25 cycles, 2 minutes at 94°C before and 7 minutes at 72°C after the PCR reaction. 0.5 μ l of the domain exchange variants was added per cycle at 50°C.

For extension at 55°C: 1 minute 95°C / 30 seconds 50°C / 1 minute 55°C // 25 cycles, 2 minutes at 95°C before and 7 minutes at 55°C after the PCR reaction. 0.5 μ l of the domain exchange variants was added per cycle at 50°C.

Example 6: 3'-5' exonuclease test - TaqEc1 variant

The samples were incubated in the absence of nucleotides with a 5'-Dig-labelled primer which anneals to a DNA template strand. 10 μ l test mix contained 1 μ l buffer (100 mM Tris-HCl, 15 mM MgCl₂, 500 mM KCl, 0.1 mg/ml gelatin, pH 8.3), 1 μ l enzyme TaqEc1 (500 units/ μ l), 1 pmol template strand (50-mer, see scheme) and 500 fmol 5'-Dig-labelled primer P1 (matched, 23mer, see scheme) or P2 (mismatched, 23mer, see scheme). The reaction mixtures were incubated at 50°C for various incubation periods. The DNA fragments were separated on a 12.5 % acrylamide gel (SequaGel Kit, Medco Company) and transferred onto a nylon membrane (Boehringer Mannheim) by contact blotting. The nylon membrane was treated as follows: 100 ml buffer 1 (1 % blocking reagent (Boehringer Mannheim) in 0.1 M maleic acid, 0.15 M NaCl, pH 7.5), incubate for 30 minutes; 100 ml buffer 2 (1:10000 diluted anti-Dig-AP Fab fragment antibody (Boehringer Mannheim) in buffer 1), incubate for 30 minutes; 135 ml buffer 3 each time (buffer 1 containing 0.3 % Tween 20), wash three times for 30 minutes; 50 ml buffer 4 (0.1 M Tris-HCl, 0.1 M NaCl, 50 mM MgCl₂, pH 9.5), incubate for 5 minutes; 50 ml buffer 5 (1:1000 diluted CPD star (Boehringer Mannheim) in buffer 4), incubate for 5 minutes. The nylon membrane was dried on Watman paper and exposed for 30 to 60 minutes on a chemiluminescence film (Boehringer Mannheim) for the chemiluminescence detection. If a 3'-5' exonuclease is present, the degradation of the primer at the 3' end is visible (see figures). The Taq polymerase with a His tag (NHis-TaqPol) was used as a negative control and the UITma DNA polymerase was used as a positive control. For both control enzymes the reactions mixtures were incubated at 72°C. The reaction buffer of the

manufacturer was used for the UITma DNA polymerase. Fig. 12 and 13 show the 3'-5' exonuclease test variant TaqEc1.

Example 7: Correction of 3'-mismatched primers and their extension - TaqEc1 variant (3'-mismatch primer correction assay)

Dig-labelled primers which anneal to a template strand (50 mer, see scheme) were extended in four different experiments. The primers were a matched primer (P1, 23mer, see scheme) and two different mismatched primers (P2, P3, 23mers, see scheme) which anneal in the recognition sequence of the restriction enzyme EcoRI. A 20 μ l test mix contained 1 μ l buffer (100 mM Tris-HCl, 15 mM MgCl₂, 500 mM KCl, 0.1 mg/ml gélatin, pH 8.3), 1 μ l enzyme TaqEc1 (500 units/ μ l), 10 μ M each of dATP, dCTP, dGTP, dTTP, 1 pmol template strand and 500 fmol of each 5'-Dig-labelled primer P1 (matched) or P2 (mismatched) or P3 (mismatched). The reaction mixtures were incubated for 60 minutes at 50°C and afterwards heated for 5 minutes to 95°C. 10 μ l aliquots were removed and cleaved for 30 minutes at 37°C with 10 units EcoRI. The DNA fragments were separated on a 12.5 % acrylamide gel (SequaGel Kit, Medco Company) and transferred by contact blot onto a nylon membrane (Boehringer Mannheim). The nylon membrane was treated as described above and exposed for 30 to 60 minutes on a chemiluminescence film (Boehringer Mannheim). When the matched primer was used, the digestion with EcoRI resulted in a 28 bp and a 18 bp fragment. The mismatched primers yield this result only when mismatched nucleotides are replaced by matched nucleotides (see Figure 14).

Example 8: Modification of a recombinant DNA polymerase Design of the hybrid polymerase gene Ath Pol and Tne Pol

Computer prediction

The structure of the chimeric polymerase gene was derived from the sequence alignment (Thompson, J.D. Higgins, D.G. and Gibson, T.J. Nucleic Acids Research, 1994, 22: 4673-4680) between the polymerases and the E. coli POLI gene - the sequence with the highest correspondence with the resolved 3D structure in the data bank of Brookhaven (for the Klenow fragment). The pair alignments showed a correspondence of ca. 40 % and hence the 1KLN structure can presumably be regarded as the best possible prototype. In order to ensure a smooth transition from one structure to the other, the crossing point should be located in an area which has a high similarity with all three proteins from the point of view of multiple alignments. The crossing point should therefore be between the polymerase and 3'-5' exonuclease domain (Figure 17, 18).

Construction of a hybrid polymerase gene and expression vectors.

Computer predictions and simulations serve as a basis for the construction of a hybrid gene. PCR amplification and subcloning were used as methods to obtain the ATH POL and TNE EXO domains in which two primer pairs having the structure shown in Fig. 18 were used. The primers have sequences which are specific for the N- and C-ends of the respective genes and for the connecting sequence in the middle of the gene as shown in Fig. 2B, C. The overlap of 12 bases in the ATHUP and TNELOW primers was designed for the subsequent reconstruction of the hybrid

gene and, furthermore, inserted in an unequivocal SalI restriction site which can be used for further modifications with polymerase domains. The overhangs of the 5' sequence of the TNEUP and ATHLOW primers code for the restriction sites NcoI and HindIII for the later subcloning of the required fragments in the expression vector.

However, applying this strategy requires extensive sequencing of the subcloned regions. For this reason an additional construct was built and the splice connection between the genes was moved to another position i.e. 42 amino acids further below the original connecting position to a region between the polymerases which has a higher similarity. An advantage of the new design is the unequivocal BamHI sequence within the TNE polymerase sequence containing the proposed splice connection. In order to construct the hybrid gene, a BamHI sequence was incorporated into the ATH polymerase sequence which is subsequently used to assemble parts of the gene by a directed mutagenesis. The amino acids and nucleotide sequence of the new compound is shown in Fig. 19.

The hybrid polymerase gene was constructed as described in Fig. 20 by multiple subcloning, directed mutagenesis and sequencing steps.

All fragments obtained by the PCR amplification were sequenced starting at the ends up to the unequivocal restriction sites used in the subsequent subcloning steps. In order to ensure the accuracy of the amplification the PCR reactions were carried out with Vent polymerase (New England Biolabs). The directed mutagenesis was carried out using the "Quick Change"

method (Stratagene).

Example 9: Expression of a hybrid polymerase gene in E. coli

The plasmids pETHYBR and pETHYBRd5 were transformed in the E. coli strain BL21 (DE3) plySS from Novagene and led to the expression of T7 polymerase.

The expression of the hybrid POL gene was monitored in the extracts of recombinant strains by measuring the DNA polymerase activity using the activated DNA assay. The following conditions were used.

- 1) The recombinant E. coli strains were cultured in LB medium containing 100 mcg/ml ampicillin + 30 mcg/ml chloroamphenicol (for pETHYBR and pETHYBRd5 in BL21 (DE3) plySS) or in 20 ml LB medium containing 100 mcg/ml ampicillin + 30 mcg/ml kanamycin (pARHYBd5 in JM109/pSB1611).
- 2) The cultures were shaken at 37°C to an optical density of OD 550 ~ 0.6-0.7; then the cultures were cooled to 25° - 28°C, IPTG was added to a final concentration of 1 mM. The incubation was then continued at 25-30°C: For two pET vectors the density of non-induced cultures after 4 hours incubation was OD 550 ~ 2.2 and for induced cultures ~ 1.5.
- 3) Protein extracts of BL21 (DE3) plySS strains were produced by pelleting 5 ml aliquots of the cultures; the cell pellets were then resuspended in 100 µl termination buffer containing 40 mM Tris-HCl, pH 8.0,

0.1 mM EDTA, 7 mM 2-mercaptoethanol, 0.2 mM PMSF, 0.1 % Triton X-100. The cell extracts were prepared by freezing and thawing the cell suspension in two cycles in liquid nitrogen/warm water bath; then a KCl solution was added to a final concentration of 0.75 M and the extracts of the induced and non-induced cultures were heated for 15 min at 72°C, pelleted and used to measure the polymerase activity; this was carried out in an activated DNA assay (100 mcg/ml activated DNA, 3 mM MgSO₄, 50 mM Tris-HCl, pH 8.9, 0.1 % Triton X-100, 70 µM dA-P33, 5-10 µCi/ml) in a volume of 20 µl using 2 µl heated cell extracts.

The results are shown in the following table:

Relative DNA polymerase activity in extracts of recombinant strains (% incorporation of labels, mean of 3 independent measurements)

| Strain | BL21 (DE3) plyS | | | |
|-------------------|-----------------|----|-----------|----|
| plasmid | pETHYBR | | pETHYBRd5 | |
| IPTG | - | + | - | + |
| TCA insoluble r/a | 5 | 40 | 2 | 85 |

These data show that both versions of the hybrid polymerase gene could be expressed with the pET vector system.

Characterization of the recombinant hybrid polymerase.

Thermal stability

The thermal stability of recombinant polymerases was determined by heating the extract of the E. coli strain for various periods (10, 30, 60, 120 minutes) at 95°C. It turned out that the completely formed as well as the shortened hybrid polymerase were not sufficiently stable (100 % inactive after a 10 minute incubation at 95°C). The degree of expression of the recombinant polymerases was evaluated by analysing the heated cell extracts in 10 % SDS PAAG; since no visible difference was found between the induced and non-induced cultures, it may be concluded that the production of hybrid polymerases does not exceed 1 % of the total soluble protein.

Proof-reading activity

The proof-reading activity of the recombinant DNA polymerase derived from pETHYBRd5, e.g. Klenow fragment was tested according to the same protocol which was also used for the archaeal DNA. It turned out that the recombinant enzyme has proof reading activity.

Reverse transcriptase activity

The following reaction mixture was used to determine the reverse transcriptase activity: 1µg polydA-(dT)₁₅, 330 µM TTP, 0.36 µM digoxigenin-dUTP, 200 µg/ml BSA, 10 mM Tris HCl, pH 8.5, 20 mM KCl. The concentration of MgCl₂ in the reaction mixture varied between 0.5 and 10 mM. DTE was added at a concentration of 10 mM.

2 μ l recombinant DNA polymerase (derived from pETHYBRd5, e.g. Klenow fragment) was added to the reaction mixture and incubated for 15 min at 50°C. Tth DNA polymerase containing Mn^{2+} was added as a positive control. After stopping the reaction, the mixture was applied to a positively charged nylon membrane (BM). The incorporated digoxigenin was detected by means of the BM protocol, 1995.

It turned out that the recombinant enzymes (Klenow fragment) have reverse transcriptase activity (Fig. 22). The activity is dependent on the presence of Mn^{2+} (optimal concentration 1 mM). The presence of Mg^{2+} had moreover an additional stimulating effect (optimal Mg^{2+} concentration 4 mM).

Example 11: Construction of the chimeric polymerase gene (see Fig. 20)

Abbreviations for the restriction sequences - B-BamHI, Bsp-BspHI, H-HindIII, N-NcoI, R-EcoRI, S-SalI, Sn-SnaI, X-XhoI, Xm-XmaI

1. PCR amplification of the ATH POL domain using the primers ATH UP and ATHLOW and the pARHis10 plasmid containing the complete polymerase gene in the vector pTrcHISB and subcloning in the pSK+Bluescript plasmid \rightarrow pBSAT. The insertion was sequenced from the flanking primers and it turned out that due to an error during the primer synthesis, a single base in the ATHUP primer sequence had been deleted.
2. Directed mutagenesis of the plasmid pARHis10 with primers m1 and m2 using the "Quick change" procedure

(Stratagene) to incorporate the BamHI sequence at position 1535 → pARHis10mut.

3. PCR amplification of the TNE EXO domain using the primers TNEUP and TNELOW on the template of the pTNEC2 plasmid and subcloning in the SmaI cut puC19 plasmid → pTEX1 and pTEX2 with different orientations of the incorporation.
3. Subcloning the 1444 bp XhoI-BamHI fragment from the pTNEC2 plasmid containing the "LONG" EXO domain in the XhoI-BamHI cut plasmid pTEX1 → pTEXL.
5. Incorporation of the complete ATH polymerase gene as a 2553 bp BamHI-HindII fragment in BamHI-HindIII cut pTEXL → pTEXLATF.
6. Substitution of the XmaI-SnaI fragment of the pTEXLATF plasmid by the 1094 bp XmaI-SnaI fragment from the pARHis10mut plasmid containing the incorporated BamHI sequence → pTEXLATF*.
7. Incorporation of the 4214 bp NcoIHindII fragment from pTEXLATF* into the NcoI-HindII cut pET21d vector → pETNAT.
8. Deletion of the 1535 bp BamHI fragment coding for the N-terminal domain of the ATH polymerase from the pETNAT plasmid; this leads to an in-frame joining of the TNE EXOL and ATH POL sequences → pETHYBR.

9. Substitution of the 1661 bp NcoI-BamHI fragment of pETHYBR by the 829 bp BspHI-BamHI fragment from pETNAT; this leads to the use of Met284 of the TNE polymerase as the starting codon and to deletion of the N-terminal domain with the assumed 5'-3' exonuclease activity → pETHYBRd5.

09/623326

534 Rec'd PCT/PTO 30 AUG2000

Claims

1. Polymerase chimera composed of functional amino acid fragments of at least two different polymerases, wherein the functional amino acid fragments are active in the polymerase chimera and the domain having polymerase activity is derived from the first polymerase and the domain having 3'-5' exonuclease activity is derived from the second polymerase and wherein the amino acid sequence of the polymerase chimera essentially corresponds to SEQ ID NO:8.
2. Polymerase chimera composed of functional amino acid fragments of at least two different polymerases, wherein the functional amino acid fragments are active in the polymerase chimera and the domain having polymerase activity is derived from the first polymerase and the domain having 3'-5' exonuclease activity is derived from the second polymerase and wherein the amino acid sequence of the polymerase chimera essentially corresponds to SEQ ID NO:10.
3. Polymerase chimera composed of functional amino acid fragments of at least two different polymerases, wherein the functional amino acid fragments are active in the polymerase chimera and the domain having polymerase activity is derived from the first polymerase and the domain having 3'-5' exonuclease activity is derived from the second polymerase and wherein the amino acid sequence of the polymerase chimera essentially corresponds to SEQ ID NO:12.

4. Polymerase chimera as claimed in one of the claims 1-3, wherein the chimera additionally has RT activity.
5. Polymerase chimera as claimed in one of the claims 1-4, wherein histine tags have been incorporated into the amino acid sequence of the chimera.
6. DNA sequence of a polymerase chimera as claimed in one of the claims 1-5.
7. DNA sequence of a polymerase chimera according to SEQ ID NO.2.
8. DNA sequence of a polymerase chimera according to SEQ ID NO.4.
9. DNA sequence of a polymerase chimera according to SEQ ID NO.6.
10. Vector containing a DNA sequence as claimed in claims 6-9.
11. Transformed cell which contains the vector as claimed in claim 10.
12. Process for the production of the polymerase chimeras as claimed in one of the claims 1-5, wherein the process comprises the following steps:
 - designing variants with the aid of amino acid sequence alignments, of 3D models or with the aid of experimentally determined 3D structures

- production of domain exchange variants by genetic engineering
 - ligating the DNA fragments into starting vectors
 - expression of the chimeras in a host which was transformed by vectors carrying DNA fragments
 - purifying the expressed polymerase chimeras.
13. Use of the polymerase chimeras as claimed in one of the claims 1-5 for PCR.
14. Use of the polymerase chimeras as claimed in one of the claims 1-5 to sequence DNA fragments.
15. Use of the polymerase chimeras as claimed in one of the claims 1-5 for RT-PCR starting with an RNA template.
16. Kit containing a polymerase chimera as claimed in one of the claims 1-5.

Abstract

The invention concerns polymerase chimeras which are composed of amino acid fragments representing domains and which combine properties of naturally occurring polymerases that are advantageous with regard to a particular application. It has surprisingly turned out that the domains from the various enzymes are active in the chimeras and exhibit cooperative behaviour. In addition the present invention concerns a process for the production of the chimeras according to the invention and the use of these chimeras for the synthesis of nucleic acids e.g. during a polymerase chain reaction. Moreover the present invention concerns a kit which contains the polymerase chimeras according to the invention.

Figure 1/1
SEQ ID No.: 1

DNA sequence:

```

1  ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA
51  AATGAGGGGC ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG
101 TCGACGGCCA CCACCTGGCC TACCGCACCT TCCACGCCCT GAAGGGCCTC
151 ACCACCAGCC GGGGGGAGCC GGTGCAGGCG GTCTACGGCT TCGCCAAGAG
201 CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC GTGGTCTTTG
251 ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG
301 GGCCGGGCCC CCACGCCGGA GGACTTTCCC CGGCAACTCG CCCTCATCAA
351 GGAGCTGGTG GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG
401 AGGCGGACGA CGTCCTGGCC AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC
451 TACGAGGTCC GCATCCTCAC CGCCGACAAA GACCTTTACC AGCTCCTTTC
501 CGACCGCATC CACGTCCTCC ACCCCGAGGG GTACCTCATC ACCCCGGCCT
551 GGCTTTGGGA AAAGTACGGC CTGAGGCCCC ACCAGTGGGC GACTACCGG
601 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG
651 GGAGAAGACG GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC
701 TCCTCAAGAA CCTGGACCGG CTGAAGCCCC CCATCCGGGA GAAGATCCTG
751 GCCCACATGG ACGATCTGAA GCTCTCCTGG GACCTGGCCA AGGTGCGCAC
801 CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG CCCGACCGGG
851 AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC
901 GAGTTCGGCC TTCTGGAAAG CCCCTATGAC AACTACGTCA CCATCCTTGA
951 TGAAGAAACA CTGAAAGCGT GGATTGCGAA GCTGGAAAAA GCGCCGGTAT
1001 TTGCATTTGA TACCGAAACC GACAGCCTTG ATAACATCTC TGCTAACCTG
1051 GTCGGGCTTT CTTTTGCTAT CGAGCCAGGC GTAGCGGCAT ATATTCCGGT
1101 TGCTCATGAT TATCTTGATG CGCCCGATCA AATCTCTCGC GAGCGTGCAC
1151 TCGAGTTGCT AAAACCGCTG CTGGAAGATG AAAAGGCGCT GAAGGTCGGG
1201 CAAAACCTGA AATACGATCG CGGTATTCTG GCGAACTACG GCATTGAACT
1251 GCGTGGGATT GCGTTTGATA CCATGCTGGA GTCCTACATT CTCAATAGCG
1301 TTGCCGGGCG TCACGATATG GACAGCCTCG CGGAACGTTG GTTGAAGCAC
1351 AAAACCATCA CTTTTGAAGA GATTGCTGGT AAAGGCCAAA ATCAACTGAC
1401 CTTTAACCAG ATTGCCCTCG AAGAAGCCGG ACGTTACGCC GCCGAAGATG
1451 CAGATGTCAC CTTGCAGTTG CATCTGAAAA TGTGGCCGGA TCTGCAAAAA
1501 CACGAGAGGC TCCTTTGGCT TTACCGGGAG GTGGAGAGGC CCTTTCCGCG
1551 TGTCCTGGCC CACATGGAGG CCACGGGGGT GCGCCTGGAC GTGGCCTATC
1601 TCAGGGCCTT GTCCCTGGAG GTGGCCGAGG AGGTCGCCCC CCTCGAGGCC
1651 GAGGTCTTCC GCCTGGCCGG CCACCCCTTC AACCTCAACT CCCGGGACCA
1701 GCTGGAAAGG GTCCTCTTTG ACGAGCTAGG GCTTCCCGCC ATCGGCAAGA
1751 CGGAGAAGAC CGGCAAGCGC TCCACCAGCG CCGCCGTCCT GGAGGCCCTC
1801 CGCGAGGCCC ACCCCATCGT GGAGAAGATC CTGCAGTACC GGGAGCTCAC
1851 CAAGCTGAAG AGCACCTACA TTGACCCCTT GCCGGACCTC ATCCACCCCA
1901 GGACGGGCCG CCTCCACACC CGCTTCAACC AGACGGCCAC GGCCACGGGC
1951 AGGCTAAGTA GCTCCGATCC CAACCTCCAG AACATCCCCG TCCGCACCCC
2001 GCTTGGGCAG AGGATCCGCC GGGCCTTCAT CGCCGAGGAG GGGTGGCTAT
2051 TGGTGGCCCT GGACTATAGC CAGATAGAGC TCAGGGTGCT GGCCACCTC
2101 TCCGCGCAGC AGAACCTGAT CCGGGTCTTC CAGGAGGGGC GGGACATCCA
2151 CACGGAGACC GCCAGCTGGA TGTTGCGCGT CCCCCGGGAG GCCGTGGACC
2201 CCCTGATGCG CCGGGCGGCC AAGACCATCA ACTTCGGGGT CCTTACGGC
2251 ATGTCGGCCC ACCGCCTCTC CCAGGAGCTA GCCATCCCTT ACGAGGAGGC
2301 CCAGGCCTTC ATTGAGCGCT ACTTTCAGAG CTTCCCCAAG GTGCGGGCCT
2351 GGATTGAGAA GACCCTGGAG GAGGGCAGGA GGCGGGGGTA CGTGGAGACC
2401 CTCTTCGGCC GCCGCCGCTA CGTGCCAGAC CTAGAGGCCC GGGTGAAGAG
2451 CGTGCGGGAG GCGGCCGAGC GCATGGCCTT CAACATGCCC GTCCAGGGCA
2501 CCGCCGCCGA CCTCATGAAG CTGGCTATGG TGAAGCTCTT CCCCAGGCTG

```

Figure 1/2
SEQ ID No.: 1

```

2551 GAGGAAATGG GGGCCAGGAT GCTCCTTCAG GTCCACGACG AGCTGGTCCT
2601 CGAGGCCCCA AAAGAGAGGG CGGAGGCCGT GGCCCGGCTG GCCAAGGAGG
2651 TCATGGAGGG GGTGTATCCC CTGGCCGTGC CCCTGGAGGT GGAGGTGGGG
2701 ATAGGGGAGG ACTGGCTCTC CGCCAAGGAG TGA

```

SEQ ID No.: 7

amino acid sequence:

```

1 MRGSHHHHHH AADDDDKMRG MLPLFEPKGR VLLVDGHHLA YRTFHALKGL
51 TTSRGEVPQA VYGFAKSLLK ALKEDGDAVI VVFDKAPSF RHEAYGGYKA
101 GRAPTPEDFP RQLALIKELV DLLGLARLEV PGYEADDVLA SLAKKAEKEG
151 YEVRILTADK DLYQLLSMRI HVLHPEGYLI TPAWLWEKYG LRPDQWADYR
201 ALTGDESDNL PGVKGIGECT ARKLLEEWGS LEALLKNLDR LKPAIREKIL
251 AHMDDLKLSW DLAKVRTDLP LEVDFAKRRE PDRERLRAFL ERLEFGSLLH
301 EFGLLESPYD NYVTILDEET LKAWIAKLEK APVFAFDTET DSLDNISANL
351 VGLSFAIEPG VAAYIPVAHD YLDAPDQISR ERALELLKPL LEDEKALKVG
401 QNLKYDRGIL ANYGIELRGI AFDTMLESYI LNSVAGRHDM DSLAERWLKH
451 KTITFEEIAG KGKNQLTFNQ IALEEAGRYA AEDADVTLQL HLKMWPDLQK
501 HERLLWLYRE VERPLSAVLA HMEATGVRLD VAYLRALSLE VAEVVARLEA
551 EVFRLAGHPF NLNSRDQLER VLFDELGLPA IGKTEKTGKR STSAÄVLEAL
601 REAHPIVEKI LQYRELTKLK STYIDPLPDL IHPRTGRLHT RFNQATATATG
651 RLSSSDPNLQ NIPVRTPLGQ RIRRAFIAEE GWLLVALDYS QIELRVLAHL
701 SGDENLIRVF QEGRDIHTET ASWMFGVPRE AVDPLMRRAA KTINFGVLYG
751 MSAHRLSQEL AIPYEEAQAF IERYFQSFPK VRAWIEKTLE EGRRRGYVET
801 LFGRRRYVPD LEARVKSURE AAERMAFNMP VQGTAAADLMK LAMVKLFPRL
851 EEMGARMLLQ VHDELVLEAP KERAFAVARL AKEVMEGVYP LAVPLEVEVG
901 IGEDWLSAKE

```

Figure 2/1
SEQ ID No.: 2

DNA sequence:

```

1  ATGAGGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA
51 AATGAGGGGGC ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG
101 TCGACGGGCCA CCACCTGGCC TACCGCACCT TCCACGCCCT GAAGGGCCTC
151 ACCACCAGCC GGGGGGAGCC GGTGCAGGCG GTCTACGGCT TCGCCAAGAG
201 CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC GTGGTCTTTG
251 ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG
301 GGCCGGGGCC CCACGCCGGA GGA CTTTCCC CGGCAACTCG CCCTCATCAA
351 GGAGCTGGTG GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG
401 AGGCGGACGA CGTCCTGGCC AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC
451 TACGAGGTCC GCATCCTCAC CGCCGACAAA GACCTTTACC AGCTCCTTTC
501 CGACCGCATC CACGTCCTCC ACCCCGAGGG GTACCTCATC ACCCCGGCCT
551 GGCTTTGGGA AAAGTACGGC CTGAGGCCCG ACCAGTGGGC CGACTACCGG
601 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG
651 GGAGAAGACG GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC
701 TCCTCAAGAA CCTGGACCGG CTGAAGCCCG CCATCCGGGA GAAGATCCTG
751 GCCACATGG ACGATCTGAA GCTCTCCTGG GACCTGGCCA AGGTGCGCAC
801 CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG CCCGACCGGG
851 AGAGGCTTAG GGCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC
901 GAGTTCGGCC TTCTGGAAAG CCCCTATGAC AACTACGTCA CCATCCTTGA
951 TGAAGAAACA CTGAAAGCGT GGATTGCGAA GCTGGAAAAA GCGCCGGTAT
1001 TTGCATTTGA TACCGAAACC GACAGCCTTG ATAACATCTC TGCTAACCTG
1051 GTCGGGCTTT CTTTTGCTAT CGAGCCAGGC GTAGCGGCAT ATATTCCGGT
1101 TGCTCATGAT TATCTTGATG CGCCCGATCA AATCTCTCGC GAGCGTGCAC
1151 TCGAGTTGCT AAAACCGCTG CTGGAAGATG AAAAGGCGCT GAAGGTCGGG
1201 CAAAACCTGA AATACGATCG CGGTATTCTG GCGAACTACG GCATTGAACT
1251 GCGTGGGATT GCGTTTGATA CCATGCTGGA GTCCTACATT CTCAATAGCG
1301 TTGCCGGGCG TCACGATATG GACAGCCTCG CGGAACGTTG GTTGAAGCAC
1351 AAAACCATCA CTTTTGAAGA GATTGCTGGT AAAGGCAAAA ATCAACTGAC
1401 CTTTAAACAG ATTGCCCTCG AAGAAGCCGG ACGTTACGCC GCCGAAGATG
1451 CAGATGTCAC CTTGCAGTTG CATCTGAAAA TGTGGCCGGA TCTGCAAAAA
1501 CACAAAGGGC CGTTGAACGT CTTGAGAAT ATCGAAATGC CGCTGGTGCC
1551 GGTGCTTTCA CGCATTGAAC GTAACGGTGT GCGCCTGGAC GTGGCCTATC
1601 TCAGGGCCTT GTCCCTGGAG GTGGCCGAGG AGATCGCCCG CCTCGAGGCC
1651 GAGGTCTTCC GCCTGGCCGG CCACCCCTTC AACCTCAACT CCCGGGACCA
1701 GCTGGAAGG GTCTCTTTG ACGAGCTAGG GCTTCCCGCC ATCGGCAAGA
1751 CGGAGAAGAC CGGCAAGCGC TCCACCAGCG CCGCCGTCCT GGAGGCCCTC
1801 CGCGAGGCCC ACCCCATCGT GGAGAAGATC CTGCAGTACC GGGAGCTCAC
1851 CAAGCTGAAG AGCACCTACA TTGACCCCTT GCCGGACCTC ATCCACCCCA
1901 GGACGGGGCC CCTCCACACC CGCTTCAACC AGACGGCCAC GGCCACGGGC
1951 AGGCTAAGTA GTCCTGATCC CAACCTCCAG AACATCCCCG TCCGCACCCC
2001 GCTTGGGCAG AGGATCCGCC GGGCCTTCAT CGCCGAGGAG GGGTGGCTAT
2051 TGGTGGCCCT GGA CTTATAGC CAGATAGAGC TCAGGGTGCT GGCCACCTC
2101 TCCGGCGACG AGAACCTGAT CCGGCTCTC CAGGAGGGGC GGGACATCCA
2151 CACGGAGACC GCCAGCTGGA TGTTGGCGGT CCCCAGGGAG GCCGTGGACC
2201 CCCTGATGCG CCGGGCGGCC AAGACCATCA ACTTCGGGGT CCTCTACGGC
2251 ATGTCGGCCC ACCGCCTCTC CAGGAGCTA GCCATCCCTT ACGAGGAGGC
2301 CCAGGCCTTC ATTGAGCGCT ACTTTAGAG CTTCCCCAAG GTGCGGGCCT
2351 GGATTGAGAA GACCTGGAG GAGGGCAGGA GGCGGGGGTA CGTGGAGACC
2401 CTCTTCGGCC GCCGCCGCTA CGTGCCAGAC CTAGAGGCCC GGGTGAAGAG
2451 CGTGCGGGAG GCGGCCGAGC GCATGGCCTT CAACATGCCC GTCCAGGGCA
2501 CCGCCGCCGA CCTCATGAAG CTGGCTATGG TGAAGCTCTT CCCAGGCTG

```

Figure 2/2
SEQ ID No.: 2

```

2551 GAGGAAATGG GGGCCAGGAT GCTCCTTCAG GTCCACGACG AGCTGGTCCT
2601 CGAGGCCCCA AAAGAGAGGG CGGAGGCCGT GGCCCGGCTG GCCAAGGAGG
2651 TCATGGAGGG GGTGTATCCC CTGGCCGTGC CCCTGGAGGT GGAGGTGGGG
2701 ATAGGGGAGG ACTGGCTCTC CGCCAAGGAG TGA

```

SEQ ID No.: 8

amino acid sequence:

```

1  MRGSHHHHHH AADDDDKMRG MLPLFEPKGR VLLVDGHHLA YRTFHALKGL
51  TTSRGEPVQA VYGFAKSLLK ALKEDGDAVI VVFDKAPSF RHEAYGGYKA
101 GRAPTPEDFP RQLALIKELV DLLGLARLEV PGYEADDVLA SLAKKAEKEG
151 YEVRILTADK DLYQLSDRI HVLHPEGYLI TPAWLWEKYG LRPDQWADYR
201 ALTGDESDNL PGVKGIGECT ARKLEEWGS LEALLKNLDR LKPAIREKIL
251 AHMDDLKLSW DLAKVRTDLP LEVDFAKRRE PDRERLRAFL ERLEFGSLLH
301 EFGLLESPYD NYVTILDEET LKAWIAKLEK APVFAFDTET DSLDNISANL
351 VGLSFAIEPG VAAYIPVAHD YLDAPDQISR ERALELLKPL LEDEKALKVG
401 QNLKYDRGIL ANYGIELRGI AFDTMLESYI INSVAGRHDM DSLAERWLKH
451 KTITFEEIAG KGKNQLTFNQ IALEEAGRYA AEDADVTLQL HLKMWPDLOK
501 HKGPLNVFEN IEMPLVPVLS RIERNGVRLD VAYLRALSLE VAEIARLEA
551 EVFRLAGHPF NLNSRDQLER VLFDELGLPA IGKTEKTGKR STSAAVLEAL
601 REAHPIVEKI LQYRELTKLK STYIDPLPDL IHPRTGRLHT RFNQATATATG
651 RLSSSDPNLQ NIPVRTPLGQ RIRRAFIAEE GWLLVALDYS QIELRVLHL
701 SGDENLIRVF QEGRDIHTET ASWMFGVPRE AVDPLMRRAA KTINFGVLYG
751 MSAHRLSQEL AIPYEEAQAF IERYFQSFPK VRAWIEKTLE EGRRRGYVET
801 LFGRRRYVPD LEARVKSVRE AAERMAFNMP VQGTADLMK LAMVKLFPR
851 EEMGARMLLQ VHDELVLEAP KERAEEAVARL AKEVMEGVYP LAVPLEVEVG
901 IGEDWLSAKE

```

Figure 3/1
SEQ ID No.: 3

DNA sequence:

```

1  ATGAGGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA
51  AATGAGGGGGC ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG
101 TCGACGGGCA CCACCTGGCC TACCGCACCT TCCACGCCCT GAAGGGCCTC
151 ACCACCAGCC GGGGGGAGCC GGTGCAGGCG GTCTACGGCT TCGCCAAGAG
201 CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC GTGGTCTTTG
251 ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG
301 GGCCGGGGCC CCACGCCGGA GGACTTTCCC CGGCAACTCG CCCTCATCAA
351 GGAGCTGGTG GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG
401 AGGCGGACGA CGTCCTGGCC AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC
451 TACGAGGTCC GCATCCTCAC CGCCGACAAA GACCTTTACC AGCTCCTTTC
501 CGACCGCATC CACGTCTCTC ACCCCGAGGG GTACCTCATC ACCCCGGCCT
551 GGCTTTGGGA AAAGTACGGC CTGAGGCCCC ACCAGTGGGC CGACTACGGG
601 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG
651 GGAGAAGACG GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC
701 TCCTCAAGAA CCTGGACCGG CTGAAGCCCC CCATCCGGGA GAAGATCCTG
751 GCCCACATGG ACGATCTGAA GCTCTCCTGG GACCTGGCCA AGGTGCGCAC
801 CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG CCCGACCGGG
851 AGAGGCTTAG GGCCTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC
901 GAGTTCGGCC TTCTGAAAAG CCCCCCGTT GGATACAGAA TAGTGAAAGA
951 CCTGGTGAA TTTGAAAAAC TCATAGAGAA ACTGAGAGAA TCCCCTTCGT
1001 TCGCCATAGA TCTTGAGACG TCTTCCCTCG ATCCTTTCGA CTGCGACATT
1051 GTCGGTATCT CTGTGTCTTT CAAACCAAAG GAAGCGTACT ACATACCACT
1101 CCATCATAGA AACGCCCAGA ACCTGGATGA AAAAGAAGTT CTGAAAAGC
1151 TAAAAGAAAT CCTGGAGGAC CCCGGAGCAA AGATCGTTGG TCAGAATTG
1201 AAATTCGATT ACAAGGTGTT GATGGTAAAG GGTGTTGAAC CTGTCCCTCC
1251 TCACTTCGAC ACGATGATAG CGGCTTACCT TCTTGAGCCG AACGAAAAGA
1301 AGTTCAATCT GGACGATCTC GCATTGAAAT TTCTTGATA CAAAATGACC
1351 TCTTACCAGG AACTCATGTC CTTCTCTTCT CCGCTGTTTG GTTTCAGTTT
1401 TGCCGATGTT CCTGTAGAAA AAGCAGCGAA CTATTCCCTGT GAAGATGCCG
1451 ACATCACCTA CAGACTCTAC AAGATCCTGA GCTTAAAACT CCACGAGGAG
1501 AGGCTCCTTT GGCTTTACCG GGAGGTGGAG AGGCCCTTT CCGCTGTCTT
1551 GGCCACATG GAGGCCACGG GGGTGCGCCT GGACGTGGCC TATCTCAGGG
1601 CCTTGTCCCT GGAGGTGGCC GAGGAGATCG CCCGCTCGA GGCCTGAGGTC
1651 TTCCGCCTGG CCGGCCACCC CTTCAACCTC AACTCCCGGG ACCAGCTGGA
1701 AAGGGTCCTC TTTGACGAGC TAGGGCTTCC CGCCATCGGC AAGACGGAGA
1751 AGACCGCAA GCGCTCCACC AGCGCCCGCG TCCTGGAGGC CCTCCGCGAG
1801 GCCCACCCCA TCGTGGAGAA GATCCTGCAG TACCGGGAGC TCACCAAGCT
1851 GAAGAGCACC TACATTGACC CTTGCCGGA CCTCATCCAC CCCAGGACGG
1901 GCCGCCTCCA CACCCGCTTC AACCAGACGG CCACGGCCAC GGGCAGGCTA
1951 AGTAGCTCCG ATCCCAACCT CCAGAACATC CCCGTCCGCA CCCCCTTGG
2001 GCAGAGGATC CGCCGGGCTT TCATCGCCGA GGAGGGGTGG CTATTGGTGG
2051 CCCTGGACTA TAGCCAGATA GAGCTCAGGG TGCTGGCCCA CCTCTCCGGC
2101 GACGAGAACC TGATCCGGGT CTTCCAGGAG GGGCGGGACA TCCACACGGA
2151 GACCGCCAGC TGGATGTTCT GCGTCCCCCG GGAGGCCGTG GACCCCTGA
2201 TGCGCCGGGC GGCCAAGACC ATCAACTTCG GGGTCCTCTA CGGCATGTCT
2251 GCCCACCGCC TCTCCAGGA GCTAGCCATC CTTACGAGG AGGCCAGGC
2301 CTTCAATTGAG CGCTACTTTC AGAGCTTCCC CAAGGTGCGG GCCTGGATTG
2351 AGAAGACCTT GGAGGAGGGC AGGAGGCGGG GGTACGTGGA GACCCTCTTC
2401 GGCCGCGGCC GCTACGTGCC AGACCTAGAG GCCCGGTGA AGAGCGTGCG
2451 GGAGGCGGCC GAGCGCATGG CTTCAACAT GCCCGTCCAG GGCACCGCCG
2501 CCGACCTCAT GAAGCTGGCT ATGGTGAAGC TCTTCCCCAG GCTGGAGGAA

```

Figure 3/2
SEQ ID No.: 3

```

2551 ATGGGGGCCA GGATGCTCCT TCAGGTCCAC GACGAGCTGG TCCTCGAGGC
2601 CCCAAAAGAG AGGGCGGAGG CCGTGGCCCG GCTGGCCAAG GAGGTCATGG
2651 AGGGGGTGTA TCCCCTGGCC GTGCCCCTGG AGGTGGAGGT GGGGATAGGG
2701 GAGGACTGGC TCTCCGCCAA GGAGTGA

```

SEQ ID No.: 9

amino acid sequence:

```

1   MRGSHHHHHH AADDDDKMRG MLPLFEPKGR VLLVDGHHLA YRTFHALKGL
51  TTSRGEPVQA VYGFAKSLLK ALKEDGDAVI VVFDKAPSF RHEAYGGYKA
101 GRAPTPEDFP RQLALIKELV DLLGLARLEV PGYEADDVLA SLAKKAEKEG
151 YEVRIILTADK DLYQLLSDRI HVLHPEGYLI TPAWLWEKYG LRPDQWADYR
201 ALTGDESDNL PGVKGIGECT ARKLLEWGS LEALLKNLDR LKPAIREKIL
251 AHMDDLKLSW DLAKVRTDLP LEVDFAKRRE PDRERLRAFL ERLEFGSLH
301 EFGLLSPPV GYRIVKDLVE FEKIEKLRE SPSFAIDLET SSLDPFDCDI
351 VGISVSFKPK EAYYIPLHHR NAQNLDEKEV LKKLKEILED PGAKIVGQNL
401 KFDYKVLNVK GVEPVPPHFD TMIAAYLLEP NEKKFNLDDL ALKFLGYKMT
451 SYQELMSFSS PLFGFSFADV PVEKAANYSC EDADITYRLY KILSLKLHEE
501 RLLWLYREVE RPLSAVLAHM EATGVRLDVA YLRALSLEVA EEIARLEAEV
551 FRLAGHPFNL NSRDQLERVL FDELGLPAIG KTEKTGKRST SAAVLEALRE
601 AHPIVEKILQ YRELTKLKST YIDPLDLIH PRTGRLHTRF NQTATATGRL
651 SSSDPNLQNI PVRTPLGQRI RRAFIAEEGW LLVALDYSQI ELRVLAHLG
701 DENLIRVFQE GRDIHTETAS WMFGVPREAV DPLMRRAAKT INFGVLYGMS
751 AHRLSQELAI PYEEAQAFIE RYFQSFPKVR AWIEKTLEEG RRRGYVETLF
801 GRRRYVPDLE ARVKSUREAA ERMAFNMPVQ GTAADLMKLA MVKLFPRLEE
851 MGARMLLQVH DELVLEAPKE RAEAVARLAK EVMEGVYPLA VPLEVEVGIG
901 EDWLSAKE

```

Figure 4/1
SEQ ID No.: 4

DNA sequence:

```

1  ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA
51  AATGAGGGGC ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG
101 TCGACGGCCA CCACCTGGCC TACCGCACCT TCCACGCCCT GAAGGGCCTC
151 ACCACCAGCC GGGGGGAGCC GGTGCAGGCG GTCTACGGCT TCGCCAAGAG
201 CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC GTGGTCTTTG
251 ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG
301 GGCCGGGCCC CCACGCCGGA GGACTTTCCC CGGCAACTCG CCCTCATCAA
351 GGAGCTGGTG GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG
401 AGGCGGACGA CGTCCTGGCC AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC
451 TACGAGGTCC GCATCCTCAC CGCCGACAAA GACCTTTACC AGCTCCTTTC
501 CGACCGCATC CACGTCTCC ACCCGAGGG GTACCTCATC ACCCCGGCCT
551 GGCTTTGGGA AAAGTACGGC CTGAGGCCCG ACCAGTGGGC CGACTACCGG
601 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG
651 GGAGAAGACG GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC
701 TCCTCAAGAA CCTGGACCGG CTGAAGCCCG CCATCCGGGA GAAGATCCTG
751 GCCCACATGG ACGATCTGAA GCTCTCCTGG GACCTGGCCA AGGTGCGCAC
801 CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG CCCGACCGGG
851 AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC
901 GAGTTCGGCC TTCTGGAAAG CCCCCCGTT GGATACAGAA TAGTGAAAGA
951 CCTGGTGGA TTTGAAAAAC TCATAGAGAA ACTGAGAGAA TCCCCTTCGT
1001 TCGCCATAGA TCTTGAGACG TCTTCCCTCG ATCCTTTCGA CTGCGACATT
1051 GTCGGTATCT CTGTGTCTTT CAAACCAAAG GAAGCGTACT ACATACCACT
1101 CCATCATAGA AACGCCCAGA ACCTGGATGA AAAAGAAGTT CTGAAAAAGC
1151 TAAAAGAAAT CCTGGAGGAC CCCGGAGCAA AGATCGTTGG TCAGAATTTG
1201 AAATTCGATT ACAAGGTGTT GATGGTAAAG GGTGTTGAAC CTGTCCCTCC
1251 TCACTTCGAC ACGATGATAG CGGCTTACCT TCTTGAGCCG AACGAAAAGA
1301 AGTTCAATCT GGACGATCTC GCATTGAAAT TTCTTGATA CAAAATGACC
1351 TCTTACCAGG AACTCATGTC CTTCTCTTCT CCGCTGTTTG GTTTCAGTTT
1401 TGCCGATGTT CCTGTAGAAA AAGCAGCGAA CTATTCCTGT GAAGATGCAG
1451 ACATCACCTA CAGACTCTAC AAGATCCTGA GCTTAAACT CCACGAGGCA
1501 GATCTGGAGA ACGTGTTCTA CAAGATAGAA ATGCCTCTTG TGAGCGTGCT
1551 TGCACGGATG GAACTGAACG GTGTGCGCCT GGACGTGGCC TAATCTCAGGG
1601 CCTTGTCCTT GGAGGTGGCC GAGGAGATCG CCCGCCTCGA GGCCGAGGTC
1651 TTCCGCTTGG CCGGCCACCC CTTCAACCTC AACTCCCGGG ACCAGCTGGA
1701 AAGGGTCCTC TTGACGAGC TAGGGCTTCC CGCCATCGGC AAGACGGAGA
1751 AGACCGGCAA GCGCTCTACC AGCGCCGCCG TCCTGGAGGC CCTCCGCGAG
1801 GCCACCCCA TCGTGGAGAA GATCCTGCAG TACCGGGAGC TCACCAAGCT
1851 GAAGAGCACC TACATTGACC CCTTGCCGGA CCTCATCCAC CCCAGGACGG
1901 GCCGCCTCCA CACCCGCTTC AACCAGACGG CCACGGCCAC GGGCAGGCTA
1951 AGTAGCTCCG ATCCCAACCT CCAGAACATC CCCGTCCGCA CCCCGCTTGG
2001 GCAGAGGATC CGCCGGGCCT TCATCGCCGA GGAGGGGTGG CTATTGGTGG
2051 CCCTGGACTA TAGCCAGATA GAGCTCAGGG TGCTGGCCCA CCTCTCCGGC
2101 GACGAGAACC TGATCCGGGT CTTCCAGGAG GGGCGGGACA TCCACACGGA
2151 GACCGCCAGC TGGATGTTTG GCGTCCCCCG GGAGGCCGTG GACCCCTGA
2201 TGCGCCGGGC GGCCAAGACC ATCAACTTCG GGGTCTCTA CGGCATGTG
2251 GCCCACC GCC TCTCCAGGA GCTAGCCATC CCTTACGAGG AGGCCCAGGC
2301 CTTTATTGAG CGCTACTTTC AGAGCTTCCC CAAGGTGCGG GCCTGGATTG
2351 AGAAGACCCT GGAGGAGGGC AGGAGGCGGG GGTACGTGGA GACCCTCTTC
2401 GGCCGCCGCC GCTACGTGCC AGACCTAGAG GCCCGGTGA AGAGCGTGCG
2451 GGAGGCGGCC GAGCGCATGG CCTTCAACAT GCCCGTCCAG GGCACCGCCG
2501 CCGACCTCAT GAAGCTGGCT ATGGTGAAGC TCTTCCCCAG GCTGGAGGAA

```


Figure 4/2
SEQ ID No.: 4

```

2551 ATGGGGGGCCA GGATGCTCCT TCAGGTCCAC GACGAGCTGG TCCTCGAGGC
2601 CCCAAAAGAG AGGGCGGAGG CCGTGGCCCG GCTGGCCAAG GAGGTCATGG
2651 AGGGGGTGTA TCCCTGGCC GTGCCCCTGG AGGTGGAGGT GGGGATAGGG
2701 GAGGACTGGC TCTCCGCCAA GGAGTGA

```

SEQ ID NO.: 10

amino acid sequence:

```

1  MRGSHHHHHH AADDDDKMRG MLPLFEPKGR VLLVDGHHLA YRTFHALKGL
51  TTSRGEVPQA VYGFASLLK ALKEDGDAVI VVFDKAPSF RHEAYGGYKA
101 GRAPTPEDFP RQLALIKELV DLLGLARLEV PGYEADDVLA SLAKKAEKEG
151 YEVRILTADK DLYQLLSMRI HVLHPEGYLI TPWLWEKYG LRPDQWADYR
201 ALTGDESDNL PGVKGIGECT ARKLEEWGS LEALLKNLDR LKPAIREKIL
251 AHMDDLKLSW DLAKVRTDLP LEVDFAKRRE PDRERLRAFL ERLEFGSLLH
301 EFGLESPPV GYRIVKDLVE FEKLIEKLRE SPSFAIDLET SSLDPFDCDI
351 VGISVSFKPK EAYYIPLHHR NAQNLDEKEV LKKLKEILED PGAKIVGQNL
401 KFDYKVLNVK GVEPVPPHFD TMIAAYLLEP NEKKFNLDDL ALKFLGYKMT
451 SYQELMSFSS PLFGFSFADV PVEKAANYSC EDADITYRLY KILSLKLHEA
501 DLENVFKIE MPLVSVLARM ELNGVRLDVA YLRALSLEVA EEIARLEAEV
551 FRLAGHPFNL NSRDQLERVL FDELGLPAIG KTEKTGKRST SAAVLEALRE
601 AHPIVEKILQ YRELTKLKST YIDPLPLIH PRTGRLHTRF NQTATATGRL
651 SSSDPNLQNI PVRTPLGQRI RRAFIAEEGW LLVALDYSQI ELRVLAHLG
701 DENLIRVFQE GRDIHTETAS WMFGVPREAV DPLMRRAKT INFGVLYGMS
751 AHRLSQELAI PYEEAQAFIE RYFQSFQVR AWIEKTLEEG RRRGYVETLF
801 GRRRYVPDLE ARVKSAREAA ERMAFNMPVQ GTAADLMKLA MVKLFPRLEE
851 MGARMLLQVH DELVLEAPKE RAEAVARLAK EVMEGVYPLA VPLEVEVGIG
901 EDWLSAKE

```

Figure 5/1
SEQ ID No.: 5

DNA sequence:

```

1  ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA
51  AATGAGGGGC ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG
101 TCGACGGCCA CCACCTGGCC TACCGCACCT TCCACGCCCT GAAGGGCCTC
151 ACCACCAGCC GGGGGGAGCC GGTGCAGGCG GTCTACGGCT TCGCCAAGAG
201 CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC GTGGTCTTTG
251 ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG
301 GGCCGGGGCCC CCACGCCGGA GGA CTTTCCC CGGCAACTCG CCCTCATCAA
351 GGAGCTGGTG GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG
401 AGGCGGACGA CGTCCTGGCC AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC
451 TACGAGGTCC GCATCCTCAC CGCCGACAAA GACCTTTACC AGCTCCTTTC
501 CGACCGCATC CACGTCCTCC ACCCCGAGGG GTACCTCATC ACCCCGGCCT
551 GGCTTTGGGA AAAGTACGGC CTGAGGCCCG ACCAGTGGGC CGACTACCGG
601 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGCGATCGG
651 GGAGAAGACG GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC
701 TCCTCAAGAA CCTGGACCGG CTGAAGCCCG CCATCCGGGA GAAGATCCTG
751 GCCCACATGG ACGATCTGAA GCTCTCCTGG GACCTGGCCA AGGTGCGCAC
801 CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG CCCGACCGGG
851 AGAGGCTTAG GGCCCTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC
901 GAGTTCGGCC TTCTGGAAAG CCCCCATCCA GCAGTTGTGG ACATCTTCGA
951 ATACGATATT CCATTTGCAA AGAGATACCT CATCGACAAA GGCCTAATAC
1001 CAATGGAGGG GGAAGAAGAG CTAAAGATTC TTGCCTTCGA TATAGAAACC
1051 CTCTATCACG AAGGAGAAGA GTTTGGAAAA GGCCCAATTA TAATGATTAG
1101 TTATGCAGAT GAAAATGAAG CAAAGGTGAT TACTTGGAAA AACATAGATC
1151 TTCCATACGT TGAGGTTGTA TCAAGCGAGA GAGAGATGAT AAAGAGATTT
1201 CTCAGGATTA TCAGGGAGAA GGATCCTGAC ATTATAGTTA CTTATAATGG
1251 AGACTCATTC GACTTCCCAT ATTTAGCGAA AAGGGCAGAA AAACCTGGGA
1301 TTAAATTAAC CATTGGAAGA GATGGAAGCG AGCCCAAGAT GCAGAGAATA
1351 GGCGATATGA CGGCTGTAGA AGTCAAGGGA AGAATACATT TCGACTTGTA
1401 TCATGTAATA ACAAGGACAA TAAATCTCCC AACATACACA CTAGAGGCTG
1451 TATATGAAGC AATTTTTTGA AAGCCAAAGG AGAAGGTATA CGCCGACGAG
1501 ATAGCAAAAG CCTGGGAAAG TGGAGAGAAC CTTGAGAGAG TTGCCAAATA
1551 CTCGATGGAA GATGCAAAGG CAACTTATGA ACTCGGGAAA GAATTCCTTC
1601 CAATGGAAAT TCAGCTTTCA GAGAGGCTCC TTTGGCTTTA CCGGAGGTG
1651 GAGAGGCCCC TTTCCGCTGT CCTGGCCCAC ATGGAGGCCA CGGGGGTGCG
1701 CCTGGACGTG GCCTATCTCA GGGCCTTGTC CCTGGAGGTG GCCGAGGAGA
1751 TCGCCCGCCT CGAGGCCGAG GTCTTCCGCC TGGCCGGCCA CCCCTTCAAC
1801 CTCAACTCCC GGGACCAGCT GGAAAGGGTC CTCTTTGACG AGCTAGGGCT
1851 TCCCGCCATC GGCAAGACGG AGAAGACCGG CAAGCGCTCC ACCAGCGCCG
1901 CCGTCCTGGA GGCCCTCCGC GAGGCCACC CCATCGTGGA GAAGATCCTG
1951 CAGTACCGGG AGCTACCAA GCTGAAGAGC ACCTACATTG ACCCCTTGCC
2001 GGACCTCATC CACCCAGGA CGGGCCGCCT CCACACCCGC TTCAACCAGA
2051 CGGCCACGGC CACGGGCAGG CTAAGTAGCT CCGATCCCAA CCTCCAGAAC
2101 ATCCCCGTCC GCACCCGCT TGGGCAGAGG ATCCGCCGGG CCTTCATCGC
2151 CGAGGAGGGG TGGCTATTGG TGGCCCTGGA CTATAGCCAG ATAGAGCTCA
2201 GGGTGCTGGC CCACCTCTCC GGCGACGAGA ACCTGATCCG GGTCTTCCAG
2251 GAGGGGCGGG ACATCCACAC GGAGACCGCC AGCTGGATGT TCGGCGTCCC
2301 CCGGGAGGCC GTGGACCCCC TGATGCGCCG GGCGGCCAAG ACCATCAACT
2351 TCGGGGTCTC CTACGGCATG TCGGCCACC GCCTCTCCA GGAGCTAGCC
2401 ATCCCTTACG AGGAGGCCCA GGCCTTCATT GAGCGCTACT TTCAGAGCTT
2451 CCCCAGGTG CGGGCCTGGA TTGAGAAGAC CCTGGAGGAG GGCAGGAGGC
2501 GGGGGTACGT GGAGACCCTC TTCGGCCGCC GCCGCTACGT GCCAGACCTA

```

Figure 5/2
SEQ ID No.: 5

```

2551 GAGGCCCGGG TGAAGAGCGT GCGGGAGGCG GCCGAGCGCA TGGCCTTCAA
2601 CATGCCCGTC CAGGGCACCG CCGCCGACCT CATGAAGCTG GCTATGGTGA
2651 AGCTCTTCCC CAGGCTGGAG GAAATGGGGG CCAGGATGCT CCTTCAGGTC
2701 CACGACGAGC TGGTCCTCGA GGCCCCAAAA GAGAGGGCGG AGGCCGTGGC
2751 CCGGCTGGCC AAGGAGGTCA TGGAGGGGGT GTATCCCCTG GCCGTGCCCC
2801 TGGAGGTGGA GGTGGGGATA GGGGAGGACT GGCTCTCCGC CAAGGAGTGA

```

SEQ ID No.: 11

amino acid sequence:

```

1  MRGSHHHHHH AADDDDKMRG MLPLFEPKGR VLLVDGHHLA YRTFHALKGL
51  TTSRGEVPQA VYGFAKSLLK ALKEDGDAVI VVFDKAPSF RHEAYGGYKA
101 GRAPTPEDFP RQLALIKELV DLLGLARLEV PGYEADDVLA SLAKKAEKEG
151 YEVRILTADK DLYQLLSMRI HVLHPEGYLI TPAWLWEKYG LRPDQWADYR
201 ALTGDESDNL PGVKGIGECT ARKLLEEWGS LEALLKNLDR LKPAIREKIL
251 AHMDDLKLSW DLAKVRTDLP LEVDFAKRRE PDRERLRAFL ERLEFGSLH
301 EFGLLSPPH AVVDIFEYDI PFAKRYLIDK GLIPMEGEEE LKILAFDIET
351 LYHEGEEFGK GPIIMISYAD ENEAKVITWK NIDLPYVEV SSEREMIKRF
401 LRIIREKDPD IIVTYNGDSF DFPYLAKRAE KLGIKLTIGR DGSEPKMQRI
451 GDMTAVEVKG RIHFDLYHVI TRTINLPTYT LEAVYEAIFG KPKEKVYADE
501 IAKAWESGEN LERVAKYSME DAKATYELGK EFLPMEIQLS ERLWLIVREV
551 ERPLSAVLAH MEATGVRLDV AYLRALSLEV AEEIARLEAE VFRLAGHPFN
601 LNSRDQLERV LFDELGLPAI GKTEKTGKRS TSAAVLEALR EAHPIVEKIL
651 QYRELTKLKS TYIDPLPDLI HPRTGRLHTR FNQTATATGR LSSSDPNLQN
701 IPVRTPLGQR IRRAFIAEEG WLLVALDYSQ IELRVLAHLS GDENLIRVFQ
751 EGRDIHTETA SWMFGVPREA VDPLMRRAAK TINFGVLYGM SAHRLSQELA
801 IPYEEAQAFI ERYFQSFPKV RAWIEKTLER GRRRGYVETL FGRRRYVPDL
851 EARVKSVREA AERMAFNMPV QGTAADLMKL AMVKLFPRLE EMGARMMLQV
901 HDELVLEAPK ERAEAVARLA KEVMEGVYPL AVPLEVEVGI GEDWLSAKE*

```

Figure 6/1
SEQ ID No.: 6

DNA sequence:

```

1   ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA
51  AATGAGGGGC ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCTCTCTGG
101 TCGACGGCCA CCACCTGGCC TACCGCACCT TCCACGCCCT GAAGGGCCTC
151 ACCACCAGCC GGGGGGAGCC GGTGCAGGCG GTCTACGGCT TCGCCAAGAG
201 CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC GTGGTCTTTG
251 ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG
301 GGCCGGGCCC CCACGCCGGA GGACTTTCCC CGGCAACTCG CCCTCATCAA
351 GGAGCTGGTG GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG
401 AGGCGGACGA CGTCTGGGCC AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC
451 TACGAGGTCC GCATCCTCAC CGCCGACAAA GACCTTTACC AGCTCCTTTC
501 CGACCGCATC CACGTCTCTC ACCCCGAGGG GTACCTCATC ACCCCGGCCT
551 GGCTTTGGGA AAAGTACGGC CTGAGGCCCC ACCAGTGGGC CGACTACCGG
601 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG
651 GGAGAAGACG GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC
701 TCCTCAAGAA CCTGGACCGG CTGAAGCCCC CCATCCGGGA GAAGATCCTG
751 GCCCACATGG ACGATCTGAA GCTCTCCTGG GACCTGGCCA AGGTGCGCAC
801 CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG CCCGACCGGG
851 AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC
901 GAGTTCGGCC TTCTGGAAAG CCCC GTTAGA GAACATCCAG CAGTTGTGGA
951 CATCTTCGAA TACGATATTC CATTTGCAA GAGATACCTC ATCGACAAAG
1001 GCCTAATACC AATGGAGGGG GAAGAAGAGC TAAAGATTCT TGCCTTCGAT
1051 ATAGAAACCC TCTATCACGA AGGAGAAGAG TTTGGAAAAG GCCCAATTAT
1101 AATGATTAGT TATGCAGATG AAAATGAAGC AAAGGTGATT ACTTGGAAAA
1151 ACATAGATCT TCCATACGTT GAGGTTGTAT CAAGCGAGAG AGAGATGATA
1201 AAGAGATTTT TCAGGATTAT CAGGGAGAAG GATCCTGACA TTATAGTTAC
1251 TTATAATGGA GACTCATTCG ACTTCCCAT TTTAGCGAAA AGGGCAGAAA
1301 AACTTGGGAT TAAATTAACC ATTGGAAGAG ATGGAAGCGA GCCCAAGATG
1351 CAGAGAATAG GCGATATGAC GGCTGTAGAA GTCAAGGGAA GAATACATTT
1401 CGACTTGTAT CATGTAATAA CAAGGACAAT AAATCTCCCA ACATACACAC
1451 TAGAGGCTGT ATATGAAGCA ATTTTGGAA AGCCAAAGGA GAAGGTATAC
1501 GCCGACGAGA TAGCAAAAGC CTGGGAAAGT GGAGAGAACC TTGAGAGAGT
1551 TGCCAAATAC TCGATGGAAG ATGCAAAGGC AACTTATGAA CTCGGGAAAG
1601 AATTCCTTCC AATGGAAATT CAGCTTTCAA GATTAGTTGG ACAACCTTTA
1651 TGGGATGTTT CAAGGTCAAG CACAGGGAAC CTTGTAGAGT GGTTCCTTACT
1701 TAGGAAAGCC TACGAAAGAA ACGAAGTAGC TCCAAACAAG CCAAGTGAAG
1751 AGGAGTATCA AAGAAGGCTC AGGGAGAGCT ACACAGGTGG ATTCGTGCGC
1801 CTGGACGTGG CCTATCTCAG GGCCTTGTCC CTGGAGGTGG CCGAGGAGAT
1851 CGCCCGCCTC GAGGCCGAGG TCTTCCGCCT GGCCGGCCAC CCCTTCAACC
1901 TCAACTCCCG GGACCAGCTG GAAAGGGTCC TCTTTGACGA GCTAGGGCTT
1951 CCCGCCATCG GCAAGACGGA GAAGACCGGC AAGCGCTCCA CCAGCGCCGC
2001 CGTCCTGGAG GCCCTCCGCG AGGCCACCC CATCGTGGAG AAGATCCTGC
2051 AGTACCGGGA GCTCACCAAG CTGAAGAGCA CCTACATTGA CCCCTTGCCG
2101 GACCTCATCC ACCCCAGGAC GGGCCGCCTC CACACCCGCT TCAACCAGAC
2151 GGCCACGGCC ACGGGCAGGC TAAGTAGCTC CGATCCCAAC CTCCAGAACA
2201 TCCCCGTCCG CACCCCGCTT GGGCAGAGGA TCCGCCGGGC CTTCATCGCC
2251 GAGGAGGGGT GGCTATTGGT GGCCCTGGAC TATAGCCAGA TAGAGCTCAG
2301 GGTGCTGGCC CACCTCTCCG GCGACGAGAA CCTGATCCGG GTCTTCCAGG
2351 AGGGGCGGGA CATCCACACG GAGACCGCCA GCTGGATGTT CGGCGTCCCC
2401 CGGGAGGCCG TGGACCCCTT GATGCGCCGG GCGGCCAAGA CCATCAACTT
2451 CGGGGTCTCT TACGGCATGT CGGCCACCG CCTCTCCAG GAGCTAGCCA
2501 TCCCTTACGA GGAGGCCAG GCCTTCATTG AGCGCTACTT TCAGAGCTTC

```

Figure 6/2
SEQ ID No.: 6

```

2551 CCCAAGGTGC GGGCCTGGAT TGAGAAGACC CTGGAGGAGG GCAGGAGGCG
2601 GGGGTACGTG GAGACCTCT TCGGCCGCCG CCGCTACGTG CCAGACCTAG
2651 AGGCCCCGGT GAAGAGCGTG CGGGAGGCGG CCGAGCGCAT GGCCTTCAAC
2701 ATGCCCCGTCC AGGGCACCGC CGCCGACCTC ATGAAGCTGG CTATGGTGAA
2751 GCTCTTCCCC AGGCTGGAGG AAATGGGGGC CAGGATGCTC CTTCAGGTCC
2801 ACGACGAGCT GGTCTCGAG GCCCAAAG AGAGGGCGGA GGCCGTGGCC
2851 CGGCTGGCCA AGGAGGTCAT GGAGGGGGTG TATCCCCTGG CCGTGCCCTT
2901 GGAGGTGGAG GTGGGGATAG GGGAGGACTG GCTCTCGCC AAGGAGTGA

```

SEQ ID No.: 12

amino acid sequence:

```

1  MRGSHHHHHH AADDDDKMRG MLPLFEPKGR VLLVDGHHLA YRTFHALKGL
51  TTSRGEPVQA VYGFAKSLLK ALKEDGDAVI VVFDKAPSF RHEAYGGYKA
101 GRAPTPEDFP RQLALIKELV DLLGLARLEV PGYEADDVLA SLAKKAEKEG
151 YEVRILTADK DLYQLLSDRI HVLHPEGYLI TPAWLWEKYG LRPDQWADYR
201 ALTGDESDNL PGVKGIGECT ARKLLEEWGS LEALLKNLDR LKPAIREKIL
251 AHMDDLKLSW DLAKVRTDLP LEVDFAKRRE PDRERLRAFL ERLEFGSLLH
301 EFGLLESPVR EHPAVVDIFE YDIPFAKRYL IDKGLIPMEG EEELKILAFD
351 IETLYHEGEE FGKGPIIMIS YADENEAKVI TWKNIDLPHY EVVSSSEREMI
401 KRFLRIIREK DPDIIVTYNG DSFDFFPYLAK RAEKLGIKLT IGRDGSEPKM
451 QRIGDMTAVE VKGRIHFDLY HVITRTINLP TYTLEAVYEA IFGKPKEKVY
501 ADEIAKAWES GENLERVAKY SMEDAKATYE LGKEFLPMEI QLSRLVGQPL
551 WDVSRSSSTGN LVEWFLLRKA YERNEVAPNK PSEEEYQRRL RESYTGGFVR
601 LDVAYLRALS LEVAEEIARL EAEVFRLAGH PFNLNSRDQL ERVLFDELGL
651 PAIGKTEKTG KRSTSAAVLE ALREAHPIVE KILQYRELTK LKSTYIDPLP
701 DLIHPRTGRL HTRFNQTATA TGRLLSSDPN LQNIPTVRTPL GQRIRRAFIA
751 EEGWLLVALD YSQIELRVLA HLSGDENLIR VFQEGRDIHT ETASWMFGVP
801 REAVDPLMRR AAKTINFGVL YGMSAHRSLQ ELAIPYEEAQ AFIERYFQSF
851 PKVRAWIEKT LEEGRRRGYV ETLFGRRRYV PDLEARVKSV REAAERMAFN
901 MPVQGTADL MKLAMVKLFP RLEEMGARM LQVHDELVL EAPKERAEEAVA
951 RLAKEVMIEG YPLAVPLEVE VGIGEDWLSA KE*

```

Figure 7

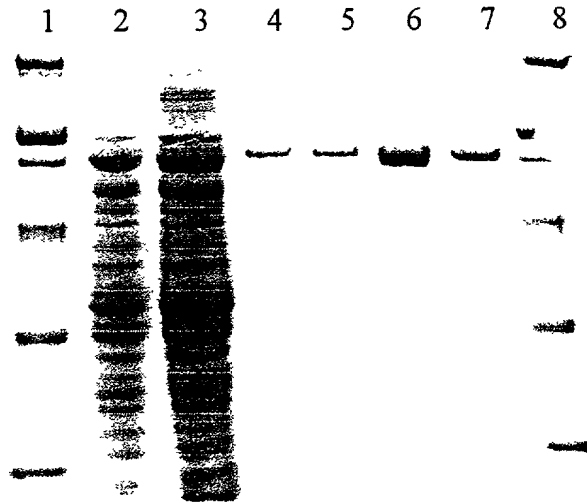


Figure 8

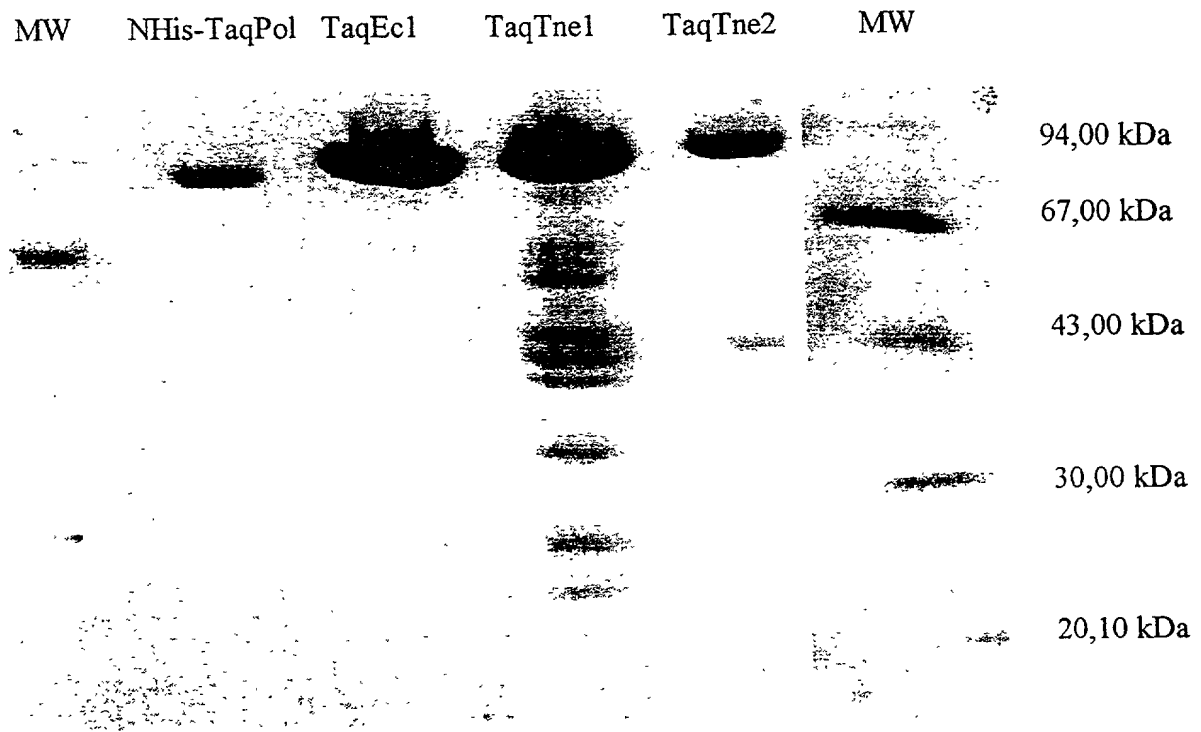


Figure 9

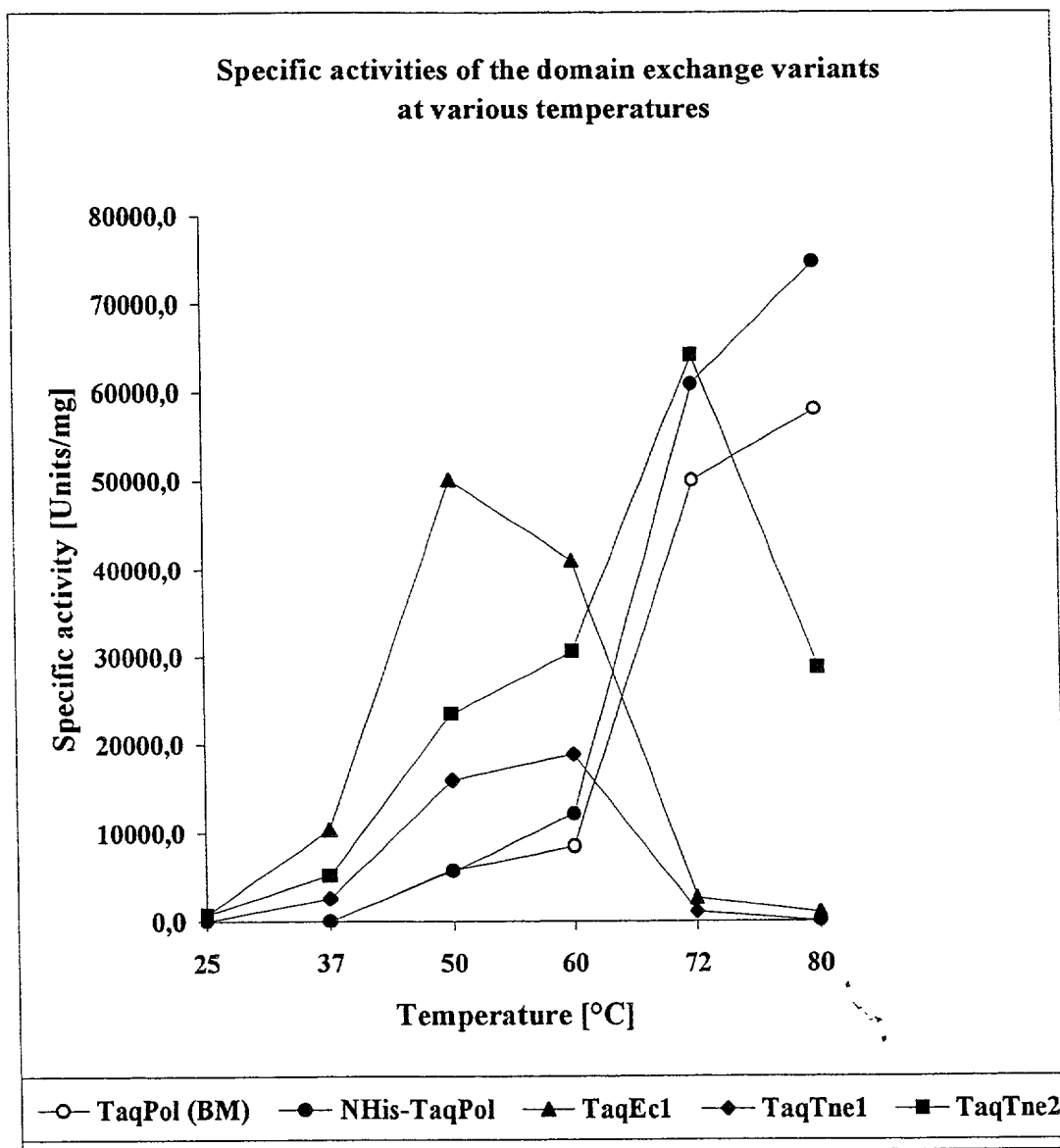


Figure 10

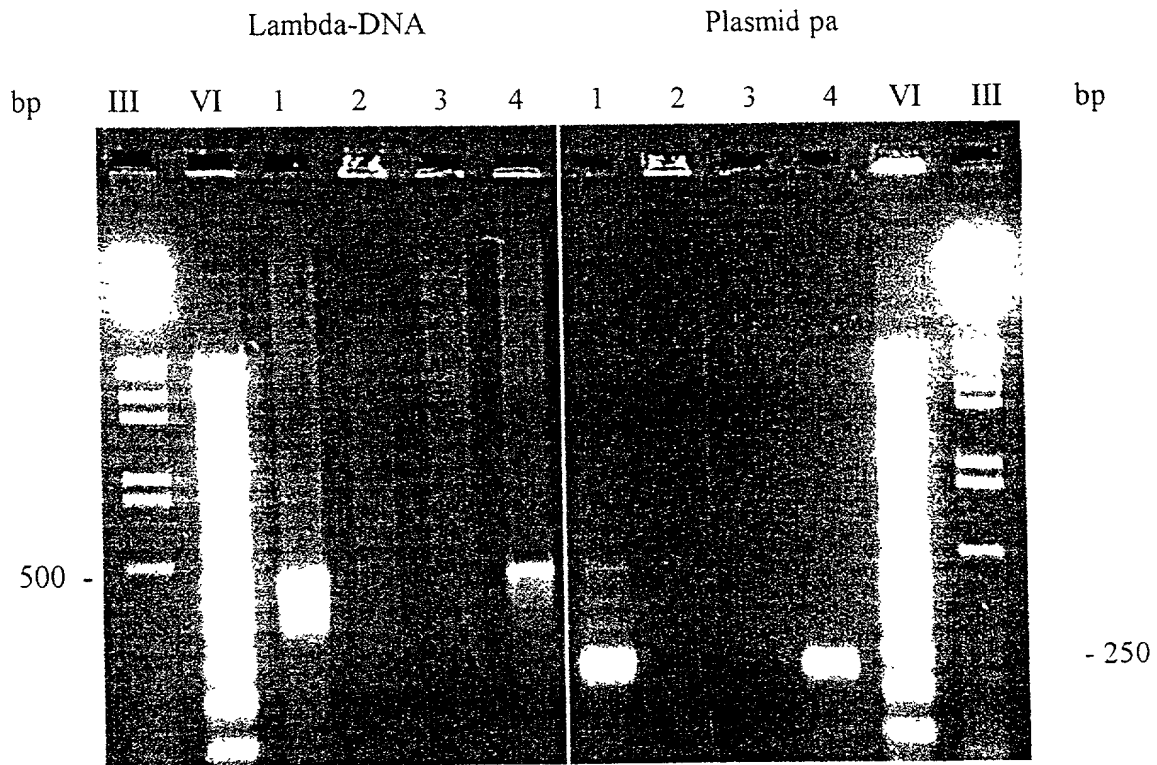


Figure 11

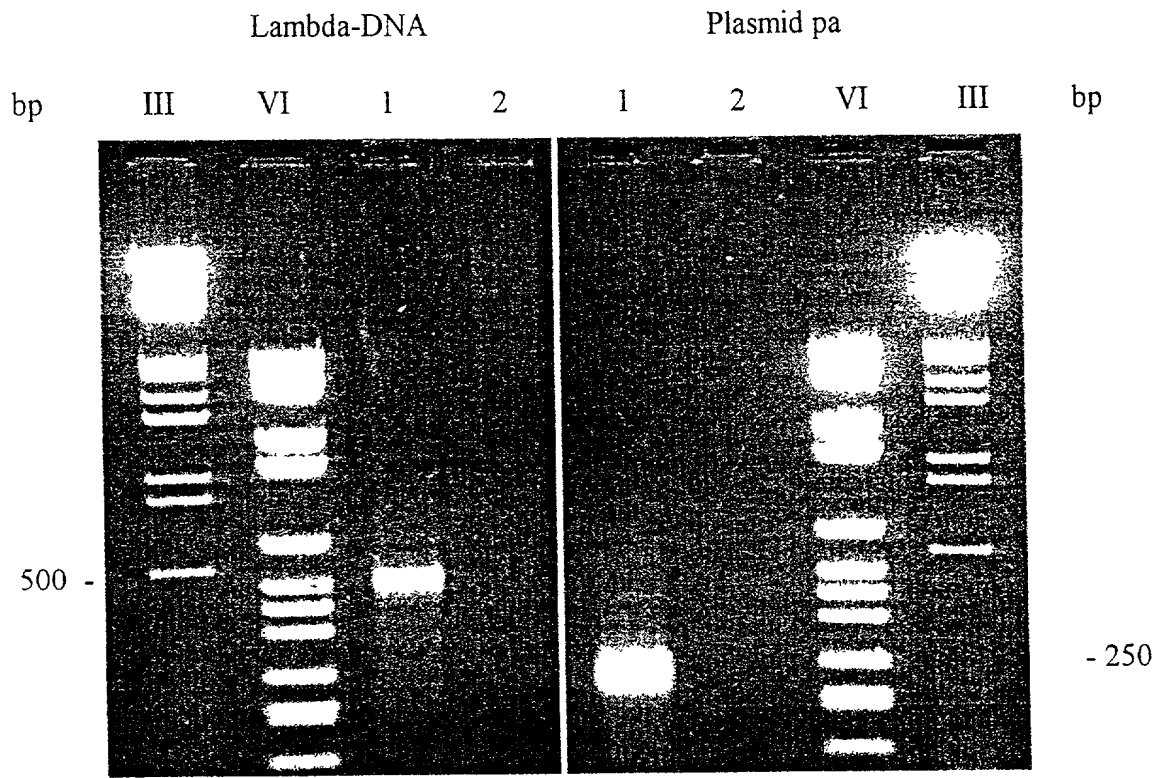


Figure 12

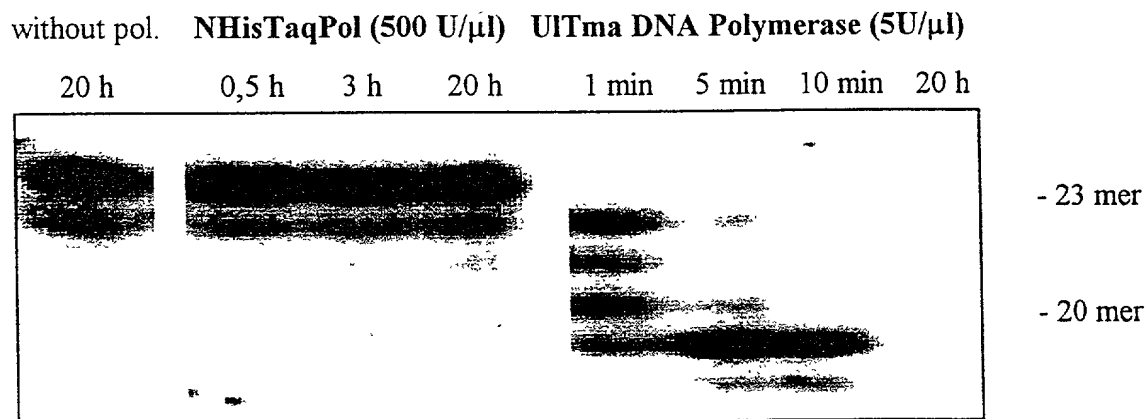
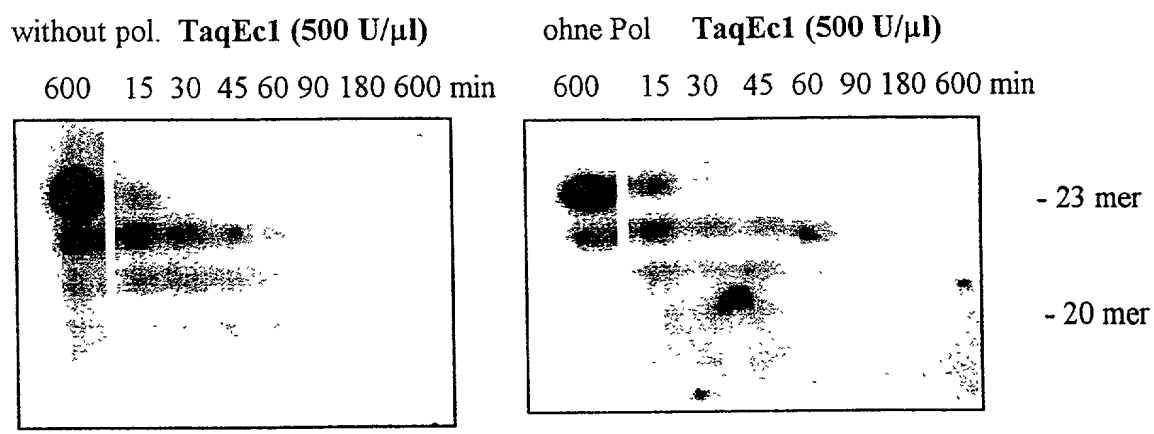


Figure 13



09/623326

Figure 14

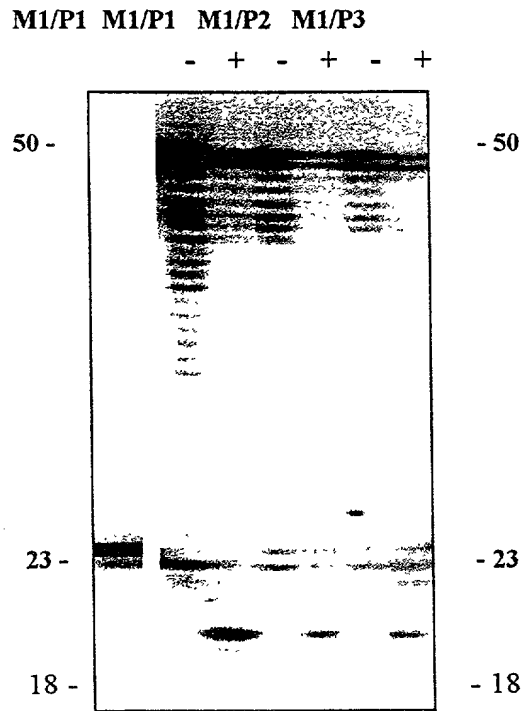


Figure 15

Degradation of primers at the 3'-end (3'-5' exonuclease assay)

and

Correction of 3'-mismatched Primers and their extension (3'-mismatch primer correction assay)

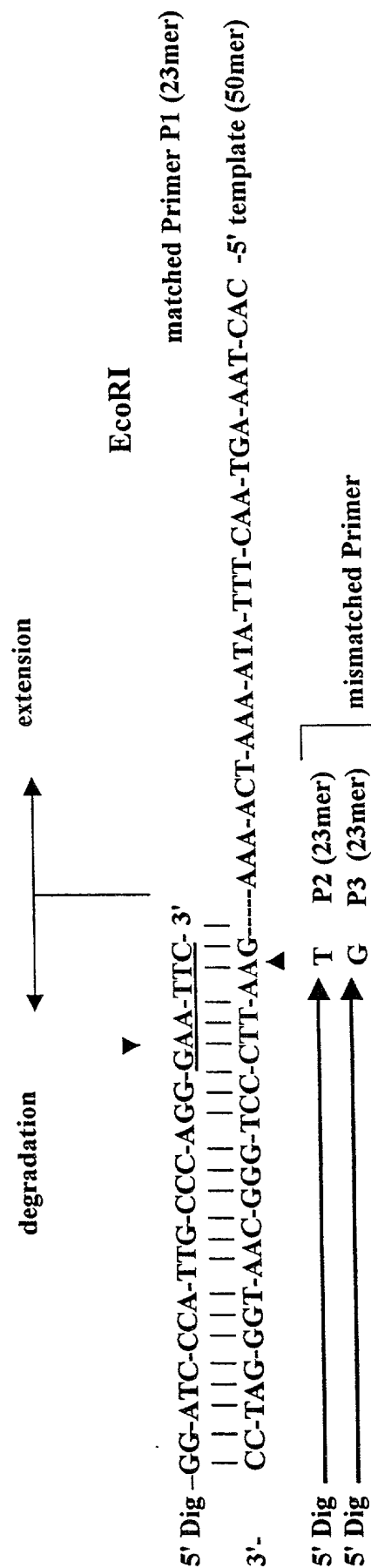


Figure 16

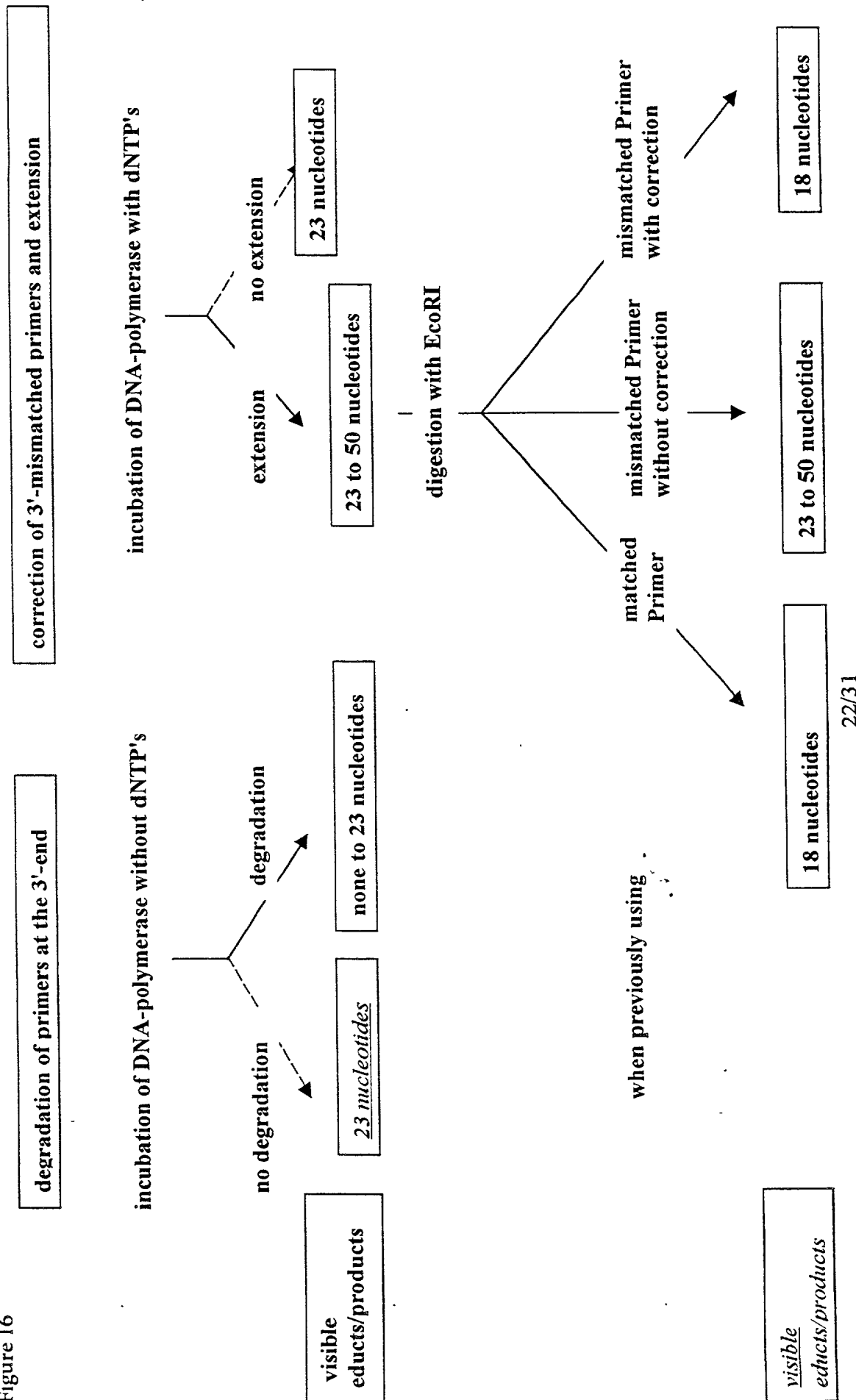


Fig. 17/1

SEQ ID No. 43
 SEQ ID No. 44
 SEQ ID No. 45
 SEQ ID No. 46

```

chimera_____-----MARLFLFDGTALAYRAYYALDRSLSTSTGIPTNATYGVARMLVRFIKDHIIIVGKD
tne.rse_____-----MARLFLFDGTALAYRAYYALDRSLSTSTGIPTNATYGVARMLVRFIKDHIIIVGKD
ath.rse_____-----MKLVIFDGN SILYRAFFALP-ELTTSNNIPTNAIYG FVN VILKYLEQ---EKPD
DPO1_ECOLI MVQIPQNPLILVDGSSYLYRAYHAFP-PLTNSAGEPTGAMYGV L NMLRSLIMQ---YKPT
               * . ** . ***. *   *. * . **.* **   ..

chimera_____YVAVAFDKKAATFRHKLLETYKAQRPKTPDLLIQQLPYIKKLVEALGMKVLEVEGYEADD
tne.rse_____YVAVAFDKKAATFRHKLLETYKAQRPKTPDLLIQQLPYIKKLVEALGMKVLEVEGYEADD
ath.rse_____YVAVAFDKRGREARKSEYEEYKANRKMPDPNLQVQIPYVREILYAFNIP IIEFEGYEADD
DPO1_ECOLI HAQVVFDAKGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMVKAMGLPLLAVSGVEADD
               ** ** .   *   * ** . *   ** * * .   . . . . * . . . .   * ****

chimera_____IIATLAVKGLPLFDEIFIVTGDKDMLQLVNEKIKVWRIVKGISD--LELYDAQKVKEKYG
chimera_____IIATLAVKGLPLFDEIFIVTGDKDMLQLVNEKIKVWRIVKGISD--LELYDAQKVKEKYG
ath.rse_____VIGSLVNQFKNTGLDIVIITGDRDTLQLLDKNVVVKIVSTKFDKTVEDLYTVENVKEKYG
DPO1_ECOLI VIGTLAREAEKAGRPVLISTGDKDMAQLVTPNITLINTMTNTILG--PE----EVVNKYG
               . * . *   .   * ****. * ** .   . .   .   * ****

chimera_____VEPQQIPDLLALTGDEIDNIPGVTGIGECTAVQLEKYKDLEDILNHVRELP-----Q
tne.rse_____VEPQQIPDLLALTGDEIDNIPGVTGIGECTAVQLEKYKDLEDILNHVRELP-----Q
ath.rse_____VWANQVPDYKALVGDQSDNIPGVKGIGEKSAQKLEEYSSLEEIYQNLDKIK-----S
DPO1_ECOLI VPPELIIDFLALMGDSSDNIPGVPGVGEKTAQALLQGLGGLDTLYAEPEKIAGLSFRGAK
               *   .   *   ** **   ***** * . ** . *   .   .   .

chimera_____KVRKALLRDRENAILS SKKLAILETNPVPIEINWEELRYQGYDREKLLPLLKELEFASIMKE
tne.rse_____KVRKALLRDRENAILS SKKLAILETNPVPIEINWEELRYQGYDREKLLPLLKELEFASIMKE
ath.rse_____SIREKLEAGKDMAFLSKRLATIVCDLPLNVKLEDLRTKEWNKERLYEILVQLEFKSIIKR
DPO1_ECOLI TMAAKLEQNKEVAYLSYQLATIKTDVELELTCEQLEVVQPPAAEELLGLFKKYEFKRWTAD
               .   *   . .   * ** . ** . . . .   * * .   * * .   **

chimera_____LQLYEESEPVGYRIVK-----DLVEFEKLIKLEKLRESP
tne.rse_____LQLYEESEPVGYRIVK-----DLVEFEKLIKLEKLRESP
ath.rse_____LGLS-----EVVQFEFVQQRDIPD
DPO1-ECOLI VEAGKWLQAKGAKPAAKPQETSVADEAPEVTATVISYDNYVTILDEETLKAWIAKLEKAP
               .

chimera_____SFAIDLETSSLDPFDCDIVGISVSFKPKKEAYYIPLHHRNAQNLDKE---VLKKLKEILE
tne.rse_____SFAIDLETSSLDPFDCDIVGISVSFKPKKEAYYIPLHHRNAQNLDKE---VLKKLKEILE
ath.rse_____VEQKELESISQIRSKE--IPLMFVQGEK-CFYLYDQESNTVFITSN-----KLLIEEIL
DPO1_ECOLI VFAFDTETDSDLNISANLVGLSFAIEPGVAAYIPVAHDYLDAPDQISRERALELLKPLLE
               . * . *   . .   . * .   .   .   *

```


Fig. 17/2

| | |
|--------------|-----------------------------------------------------------------------|
| chimera_____ | DPGAKIVGQNLKFDYKVLMMVKGVEPVPPHFDTMIAAYLLEPNEKKFNLDLALKFLGYKM |
| tne.rse_____ | DPGAKIVGQNLKFDYKVLMMVKGVEPVPPHFDTMIAAYLLEPNEKKFNLDLALKFLGYKM |
| ath.rse_____ | KSDTVKIMYDLKNI FHQLNLEDTNNIKNCEDVMIASYVLDSTRSSYELETLFVSYLNTDI |
| DPO1_ECOLI | DEKALKVGQNLKYDRGILANYGIELRGIAFDTMLESYILNSVAGRHDMSLAERWLKHKHT |
| | . . . ** * . . . * . . . * . . . * |
| chimera_____ | TSYQELMSFSSPLFGFSFADVPVEKAANYSCEDADITYRLYKILSLKLHEAD-LENV FYK |
| tne.rse_____ | TSYQELMSFSSPLFGFSFADVPVEKAANYSCEDADITYRLYKILSLKLHEAD-LENV FYK |
| ath.rse_____ | EAVKKDKKIVS-----VLLKRLWDELLRLIDLNS-CQFLYEN |
| DPO1_ECOLI | ITFEEIAGKGKNQ--LTFNQIALEEAGRYAAEDADVTQLHLKMWPDQLKHGKPLNVFEN |
| | . . . * |
| chimera_____ | IEMPLVSVLARMELNGVKVDRDALIQYTKIEINKILKLETQIYQIAGEWFNINSPKQLSY |
| tne.rse_____ | IEMPLVSVLARMELNGVYVDTEFLKKLSEEYGGKLEELAEIYRIAGEPNINSPKQVSR |
| ath.rse_____ | IERPLIPVLYEMEKTGFKVDRDALIQYTKIEINKILKLETQIYQIAGEWFNINSPKQLSY |
| DPO1-ECOLI | IEMPLVPVLSRIERNGVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQT |
| | ** ** . ** . * . * . * . . . * . . . **** ** . * ** . |
| chimera_____ | ILFEKLKLPVIKKTGTG--YSTDAEVLEELFDKHEIVPLILDYRMYTKILTTCQGLLQA |
| tne.rse_____ | ILFEKLGIKPRGKTTKTGDYSTRIEVLEELAGEHEIPLILEYRKIQKLKSTYIDALPKM |
| ath.rse_____ | ILFEKLKLPVIKKTGTG--YSTDAEVLEELFDKHEIVPLILDYRMYTKILTTCQGLLQA |
| DPO1_ECOLI | ILFEKQGIKPLKKTGPGG-APSTSEEVLEELALDYPLPKVILEYRGLAKLKSTYTDKPLPM |
| | ***** . ** . ** . ***** . . . ** . ** . * . . ** . * |
| chimera_____ | INPSSGRVHTTFIQTGTATGRLASSDPNLQNI PVKYDEGKLIRKVFVPEG-GHVLIDADY |
| tne.rse_____ | VNPKTGRIHASFNQTGTATGRLSSSDPNLQNLPTKSEEGKEIRKAIVPDPNWWIVSADY |
| ath.rse_____ | INPSSGRVHTTFIQTGTATGRLASSDPNLQNI PVKYDEGKLIRKVFVPEG-GHVLIDADY |
| DPO1-ECOLI | INPKTGRVHTSYHQAVTATGRLSSTDPNLQNI PVRNEEGRRIQAFIAPE-DYVIVSADY |
| | . ** . ** . * . . . * . ***** . * . ***** . * . . . ** . ** . . . *** |
| chimera_____ | SQIELRILAHISEDERLISAFKNNVDIHSQTAAEVFGVDIADVTPEMRSQAKAVNFGIVY |
| tne.rse_____ | SQIELRILAHLSGDENLLRAFEEGIDVHTLTASRIFNVKPEEVTEEMRRAGKMNFSIIY |
| ath.rse_____ | SQIELRILAHISEDERLISAFKNNVDIHSQTAAEVFGVDIADVTPEMRSQAKAVNFGIVY |
| DPO1_ECOLI | SQIELRIMAHLSRDKGLLTAFAGKDIHRATAAEVFGFLPLET VTSEQRRSAKAINFGLIY |
| | ***** . ** . * . * . * . * . . . * . . . ** . * . * . * . . ** . . * |
| chimera_____ | GISDYGLARDIKISRKEAAEFINKYFERYPKVKEYLDNTV KFARDNGFVLTFLNRKRYIK |
| tne.rse_____ | GVTPYGLSVRLGVPVKEAEKMIVNYFVLYPKVRDYIQRVVSEAKEKGYVRTLFGRKRDIP |
| ath.rse_____ | GISDYGLARDIKISRKEAAEFINKYFERYPKVKEYLDNTV KFARDNGFVLTFLNRKRYIK |
| DPO1_ECOLI | GMSAFGLARQLNIPRKEAQKYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLP |
| | * . . . ** *** . . . ** . ** . * . * . . . * . . . * . . . |
| chimera_____ | DIKSTNRNLRGYAERIA MN SPIQGSAADIMKLAMIKVYQKLKENNLKSKII LQVHDELLI |
| tne.rse_____ | QLMARDRNTQAEGERIAINTPIQGTAA DI IKLAMIEIDRELKERKMRSKMIIQVHDELVF |
| ath.rse_____ | DIKSTNRNLRGYAERIA MN SPIQGSAADIMKLAMIKVYQKLKENNLKSKII LQVHDELLI |
| DPO1_ECOLI | DIKSSNGARRAAAEAAINAPMQGTAA DI IKRAMIAVDLQAEQPRVRMIMQVHDELVF |
| | ** * . * . * . * . * . * * . . . * . * . * . * . |

```
chimera____ EAPYEEKDIVKEIVKREMENAVALKVPLVVEVKEGLNWYENKI
tne.rse____ EVPNEEKDALVELVKDRMTNVVKLSVPLEVDVTIGKTWS----
ath.rse____ EAPYEEKDIVKEIVKREMENAVALKVPLVVEVKEGLNWYENKI
DPO1_ECOLI EVHKDDVDAVAKQIHQLMENCTRLDVPLLVEVGSGENWDQAH-
*      .. * .      .. * *      * *** * . *      * . *
```

Figure 18/1:

SEQ ID No.: 19

SEQ ID No.: 20

SEQ ID No.: 21

SEQ ID NO.: 22

TNE UP 5' CTG ACC ATG GCG AGA CTA TTT CTC TTT G -3'
 TNE LOW 5' TCT GTC GAC CTT CAC ACC GTT CAG TTC CAT CC -3'
 ATH UP 5' - AAG GTC GAC AGA GAT GCC CTC ATC CAA TAT ACC -3'
 ATH LOW 5' - TAG CAA GCT TCT ATT TTG TCT CAT ACC AGT -3'

A.

crossing point 1

SEQ ID No.: 23

SEQ ID No.: 24

SEQ ID No.: 25

| | | | |
|------------|-------------------|--|----------------------------------------------|
| chimera__8 | IEMPLVSVLARMELNGV | | KVDRDALIQYTKIEIENKILKLETQIYQIAGEWFNINSPKQLSY |
| tne.rse__ | IEMPLVSVLARMELNGV | | YVDTEFLKKLSEEYGGKLEELAEIYRIAGEPFNINSPKQVSR |
| ath.rse__ | IERPLIPVLYEMEKTF | | KVDRDALIQYTKIEIENKILKLETQIYQIAGEWFNINSPKQLSY |

B.

SEQ ID No.: 19

SEQ ID No.: 26

SEQ ID No.: 27

5' ctg acc ATG GCG AGA CTA TTT CTC TTT G -3'
 TNEUP |----->
 ATG GCG AGA CTA TTT CTC TTT GAT GGA 27
 M A R L F L F D G 9

1

SEQ ID No.: 28

SEQ ID No.: 29

SEQ ID No.: 20

SEQ ID No.: 21

SEQ ID No.: 30

SEQ ID No.: 31

| | | |
|------|---------------------------------------------------------------------|------|
| 1512 | CGG ATG GAA CTG AAC GGT GTG TAC GTG GAC ACA GAG TTC CTG AAG AAA CTC | 1563 |
| 505 | R M E L N G V Y V D T E F L K K L | 521 |
| | 3'CC CAT CTT GAC TTG CCA CAC ctt Cag cTG TcT 5' | |

<-----| TNELOW

"Sal I site "

ATHUP |----->

5' AAG Gtc Gac AGA GAT GCC CTC ATC CAA TAT ACC -3'

| | | |
|------|---------------------------------------------------------------------|------|
| 1387 | ATG GAA AAA ACA GGA TTT AAG GTG GAT AGA GAT GCC CTC ATC CAA TAT ACC | 1435 |
| 463 | M E K T G F K V D R D A L I Q Y T | 479 |

Figure 18/2:

SEQ ID No.: 32

SEQ ID No.: 33

SEQ ID No.: 22

```
2526   GGA CTG AAC TGG TAT GAG ACA AAA TAG           2553
843     G  L  N  W  Y  E  T  K  *
      3'TG ACC ATA CTC TGT TTT ATC ttcgaacgat 5'
      <-----| ATHLOW
```

Figure 19:

SEQ ID No.: 34

SEQ ID No.: 35

SEQ ID No.: 36

crossing point 2

| | |
|------------|---------------------------------------------------------------|
| chimera__8 | IMEPLVSVLARMELNGVYVDTEFLKKLSEELYGKKLEELAEIYRIAGEPFNINSPKQVS R |
| tne.rse__ | IEMPLVSVLARMELNGVYVDTEFLKKLSEELYGKKLEELAEIYRIAGEPFNINSPKQVS R |
| ath.rse__ | IERPLIPVLYEMEKTGFKVDRDALIQYTKIENKILKLETQIYQIAGEWFFNINSPKQLS R |

A.

TNE polymerase nucleotide sequence 1642-1689

SEQ ID No.: 37

SEQ ID No.: 38

Bam HI site

=====

| | | |
|------|-----------------------------------------------------------------|------|
| 1642 | TCA CCG AAG CAG GTT TCA AGG ATC CTT TTT GAA AAA CTC GGC ATA AAA | 1689 |
| 548 | S P K Q V S R I L F E K L G I K | 563 |

SEQ ID No.: 39

SEQ ID No.: 40

SEQ ID No.: 41

SEQ ID No.: 42

ATH polymerase nucleotide sequence 1513 - 1560

| | | |
|------|-----------------------------------------------------------------|------|
| 1513 | TCA CCG AAA CAG CTT TCT TAC ATT TTG TTT GAA AAG CTA AAA CTT CCT | 1560 |
| 505 | S P K Q L S Y I L F E K L K L P | 520 |

5'CA CCG AAA CAG CTT TCT agg atc cTG TTT GAA AAG CTA AAA CTT CCT G 3'

|-----m1----->

.....3'GT GGC TTT GTC GAA AGA tcc tag gAC AAA CTT TTC GAT TTT GAA GGA C 5'

<-----m2-----|

B.

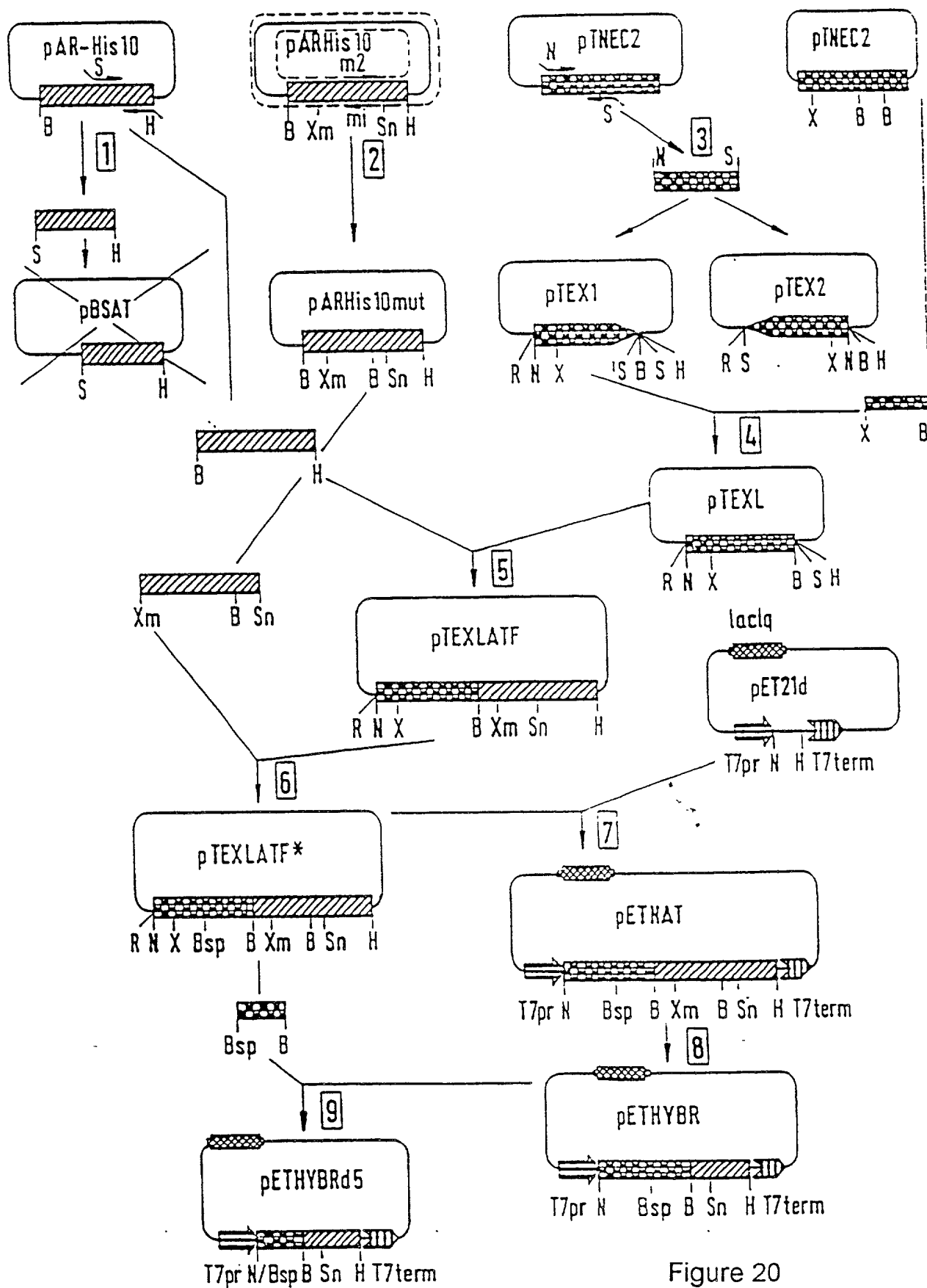


Figure 20

Figure 21

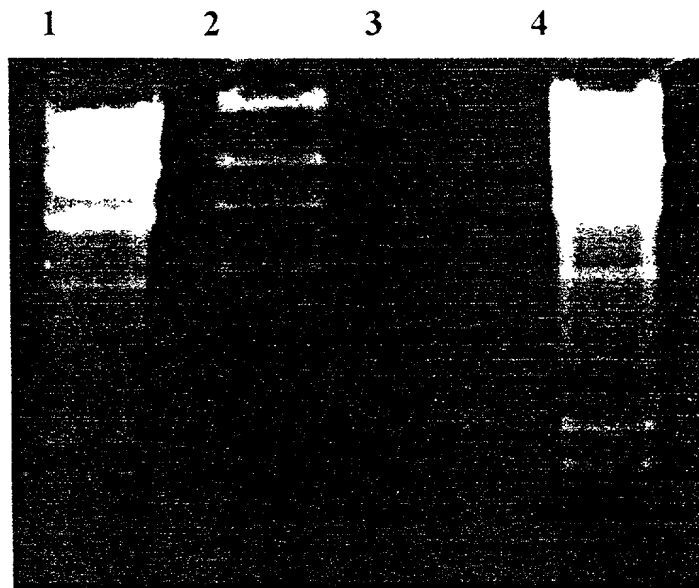
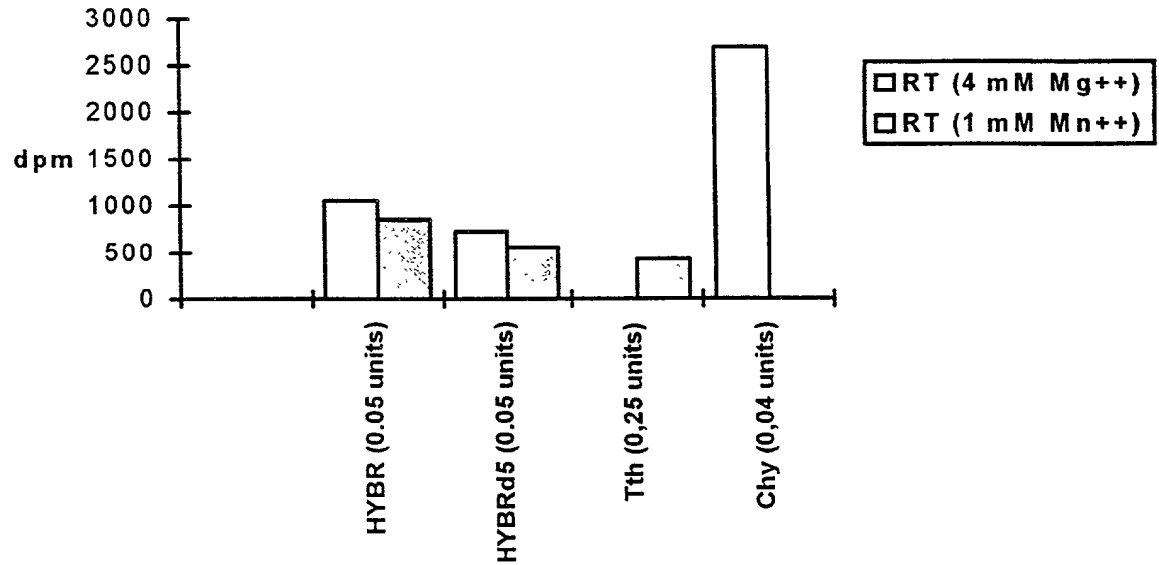


Figure 22:

Comparison of the reverse transcriptase activity of
Tne/Ath hybrid polymerases with Tth- and C.therm.
polymerase



Declaration and Power of Attorney for Patent Application

As the below named inventor(s), We hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

POLYMERASE CHIMERAS

the specification of which (check one)

☐ is attached hereto.

☐ was filed on _____ as

Application Serial No. _____

and was amended on _____ (if applicable).

☒ was filed on March 15, 1999 as

PCT International Application Serial No. PCT/EP99/01674

and was amended under PCT Article 19 on 14 April 2000 (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR §1.56.

I hereby claim foreign priority benefits under 35 U.S.C. §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Priority Claimed

| | | | | |
|---------------------------------|-----------------------------|------------------------------------------------|-----------------------------------------|-----------------------------|
| <u>198 10 879.6</u> (Number) | <u>Germany</u> (Country) | <u>13 March 1998</u> (Day/Month/Year Filed) | <input checked="" type="checkbox"/> Yes | <input type="checkbox"/> No |
| _____ (Number) | _____ (Country) | _____ (Day/Month/Year Filed) | <input type="checkbox"/> Yes | <input type="checkbox"/> No |

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below.

| | | | |
|----------------------------|------------------------|----------------------------|------------------------|
| _____ (Application No.) | _____ (Filing Date) | _____ (Application No.) | _____ (Filing Date) |
| _____ (Application No.) | _____ (Filing Date) | _____ (Application No.) | _____ (Filing Date) |

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s), or §365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

| | | |
|---------------------------------------------------|---------------------------------------|-----------------------------------------------------------|
| <u>PCT/EP99/01674</u> (Application Serial No.) | <u>15 March 1999</u> (Filing Date) | <u>Pending</u> (Status) (patented, pending, abandoned) |
| _____ (Application Serial No.) | _____ (Filing Date) | _____ (Status) (patented, pending, abandoned) |

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the practitioners at customer number 22829 to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Send Correspondence to: Customer Number 22829

Direct Telephone Calls to: Victor K. Lee, Ph.D.
(510) 814-2966

1-00

BRUNO FREY

Full Name Of Sole Or First Inventor

Bruno Frey

11/22/00

Sole Or First Inventor's Signature

Date

Penzberg, Germany

DEX

Residence

German

Citizenship

Hochfeldstrasse 50, D-82377 Penzberg, Germany

Post Office Address

2-00

BRITTA VILLBRANDT

Full Name Of Second Joint Inventor, If Any

Britta Villbrandt
Dietmar Selmer

12/18/00

Villbrandt 12/03/00

Villbrandt

Second Inventor's Signature

Date

Braunschweig, Germany

DEX

Residence

German

Citizenship

Fasanenstrasse 19, D-38102 Braunschweig, Germany

Post Office Address

30
DIETMAR SCHOMBURG

Full Name Of Third Joint Inventor, If Any

Dietmar Schomburg
Third Inventor's Signature

12/03/00
Date

Erfststadt, Germany

Residence

German

Citizenship

Richardstrasse 35, D-50374 Erfststadt, Germany

Post Office Address

40
HARALD SOBEK

Full Name Of Fourth Joint Inventor, If Any

Harald Sobek
Fourth Inventor's Signature

24. Nov. 00
Date

Penzberg, Germany

Residence

German

Citizenship

Birkenstrasse 29, D-82377 Penzberg, Germany

Post Office Address

50 WALTRAUD ANKENBAUER

Full Name Of Fifth Joint Inventor, If Any

Waltraud Ankenbauer
Fifth Inventor's Signature

23.11.00
Date

Penzberg, Germany
Residence

DEU

German
Citizenship

Oberanger 18, D-82377, Penzberg, Germany
Post Office Address

Title 37, Code of Federal Regulations, §1.56, duty to disclose information material to patentability provides, in part, that each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability as defined in this section. The duty to disclose information exists with respect to each pending claim until the claim is cancelled or withdrawn from consideration, or the application becomes abandoned.

Under this section, information is material to patentability when it is not cumulative to information already of record or being made of record in the application, and

- (1) It establishes, by itself or in combination with other information, a prima facie case of unpatentability of a claim; or
- (2) It refutes, or is inconsistent with, a position the applicant takes in:
 - (i) Opposing an argument of unpatentability relied on by the Office, or
 - (ii) Asserting an argument of patentability.

SEQUENCE PROTOCOL

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Boehringer Mannheim GmbH
(B) ROAD: Sandhoferstr. 116
(C) CITY: Mannheim
(E) COUNTRY: DE
(F) POSTAL CODE: 68305
(G) TELEPHONE: 06217595482
(H) TELEFAX: 06217594457

(ii) TITLE OF INVENTION: Polymerase chimeras

(iii) NUMBER OF SEQUENCES: 14

(iv) COMPUTER READABLE FORM:

(A) DATA CARRIER: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 2733 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single strand
(D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA AATGAGGGGC 60
ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG TCGACGGCCA CCACCTGGCC 120
TACCGCACCT TCCACGCCCT GAAGGGCCTC ACCACCAGCC GGGGGGAGCC GGTGCAGGCG 180
GTCTACGGCT TCGCCAAGAG CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC 240
GTGGTCTTTG ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG 300
GGCCGGGCCC CCACGCCGGA GGA CTTTCCC CGGCAACTCG CCCTCATCAA GGAGCTGGTG 360
GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG AGGCGGACGA CGTCCTGGCC 420
AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC TACGAGGTCC GCATCCTCAC CGCCGACAAA 480
GACCTTTACC AGCTCCTTTC CGACCGCATC CACGTCCTCC ACCCCGAGGG GTACCTCATC 540
ACCCCGGCCT GGCTTTGGGA AAAGTACGGC CTGAGGCCCG ACCAGTGGGC CGACTACCGG 600
GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG GGAGAAGACG 660
GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC TCCTCAAGAA CCTGGACCGG 720

CTGAAGCCCG CCATCCGGGA GAAGATCCTG GCCACATGG ACGATCTGAA GCTCTCCTGG 780
 GACCTGGCCA AGGTGCGCAC CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG 840
 CCCGACCGGG AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC 900
 GAGTTCGGCC TTCTGGAAAG CCCCTATGAC AACTACGTCA CCATCCTTGA TGAAGAAACA 960
 CTGAAAGCGT GGATTGCGAA GCTGGAAAAA GCGCCGGTAT TTGCATTTGA TACCGAAACC 1020
 GACAGCCTTG ATAACATCTC TGCTAACCTG GTCGGGCTTT CTTTGTCTAT CGAGCCAGGC 1080
 GTAGCGGCAT ATATTCCGGT TGCTCATGAT TATCTTGATG CGCCCGATCA AATCTCTCGC 1140
 GAGCGTGCAC TCGAGTTGCT AAAACCGCTG CTGGAAGATG AAAAGGCGCT GAAGGTCGGG 1200
 CAAAACCTGA AATACGATCG CGGTATTCTG GCGAACTACG GCATTGAACT GCGTGGGATT 1260
 GCGTTTGATA CCATGCTGGA GTCCTACATT CTCAATAGCG TTGCCGGGCG TCACGATATG 1320
 GACAGCCTCG CGGAACGTTG GTTGAAGCAC AAAACCATCA CTTTGAAGA GATTGCTGGT 1380
 AAAGGCAAAA ATCAACTGAC CTTTAACCAG ATTGCCCTCG AAGAAGCCGG ACGTTACGCC 1440
 GCCGAAGATG CAGATGTCAC CTTGCAGTTG CATCTGAAA TGTGGCCGGA TCTGCAAAAA 1500
 CACGAGAGGC TCCTTTGGCT TTACCGGGAG GTGGAGAGGC CCCTTTCCGC TGTCTGGCC 1560
 CACATGGAGG CCACGGGGGT GCGCCTGGAC GTGGCCTATC TCAGGGCCTT GTCCCTGGAG 1620
 GTGGCCGAGG AGGTCGCCCC CCTCGAGGCC GAGGTCTTCC GCCTGGCCGG CCACCCCTTC 1680
 AACCTCAACT CCCGGGACCA GCTGGAAAGG GTCCTCTTTG ACGAGCTAGG GCTTCCCGCC 1740
 ATCGGCAAGA CGGAGAAGAC CGGCAAGCGC TCCACCAGCG CCGCCGTCCT GGAGGCCCTC 1800
 CGCGAGGCC ACCCCATCGT GGAGAAGATC CTGCAGTACC GGGAGCTCAC CAAGCTGAAG 1860
 AGCACCTACA TTGACCCCTT GCCGGACCTC ATCCACCCCA GGACGGGCGG CCTCCACACC 1920
 CGCTTCAACC AGACGGCCAC GGCCACGGGC AGGCTAAGTA GCTCCGATCC CAACCTCCAG 1980
 AACATCCCCG TCCGCACCCC GCTTGGGCAG AGGATCCGCC GGGCCTTCAT CGCCGAGGAG 2040
 GGGTGGCTAT TGGTGGCCCT GGA CTATAGC CAGATAGAGC TCAGGGTGCT GGCCACCTC 2100
 TCCGGCGACG AGAACCTGAT CCGGGTCTTC CAGGAGGGGC GGGACATCCA CACGGAGACC 2160
 GCCAGCTGGA TGTTGGGCGT CCCCCGGGAG GCCGTGGACC CCCTGATGCG CCGGGCGGCC 2220
 AAGACCATCA ACTTCGGGGT CCTCTACGGC ATGTCGGCCC ACCGCCTCTC CCAGGAGCTA 2280
 GCCATCCCTT ACGAGGAGGC CCAGGCCTTC ATTGAGCGCT ACTTTCAGAG CTTCCCAAG 2340
 GTGCGGGCCT GGATTGAGAA GACCCTGGAG GAGGGCAGGA GGCGGGGGTA CGTGGAGACC 2400
 CTCTTCGGCC GCCGCCGCTA CGTGCCAGAC CTAGAGGCC GGGTGAAGAG CGTGCGGGAG 2460
 GCGGCCGAGC GCATGGCCTT CAACATGCCC GTCCAGGGCA CCGCCGCCGA CCTCATGAAG 2520

CTGGCTATGG TGAAGCTCTT CCCCAGGCTG GAGGAAATGG GGGCCAGGAT GCTCCTTCAG 2580
 GTCCACGACG AGCTGGTCCT CGAGGCCCCA AAAGAGAGGG CGGAGGCCGT GGCCCGGCTG 2640
 GCCAAGGAGG TCATGGAGGG GGTGTATCCC CTGGCCGTGC CCCTGGAGGT GGAGGTGGGG 2700
 ATAGGGGAGG ACTGGCTCTC CGCCAAGGAG TGA 2733

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS

- (A) LENGTH: 2733 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA AATGAGGGGC 60
 ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG TCGACGGCCA CCACCTGGCC 120
 TACCGCACCT TCCACGCCCT GAAGGGCCTC ACCACCAGCC GGGGGGAGCC GGTGCAGGCG 180
 GTCTACGGCT TCGCCAAGAG CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC 240
 GTGGTCTTTG ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG 300
 GGCCGGGGCC CCACGCCGGA GGA CTTTCCC CGGCAACTCG CCCTCATCAA GGAGCTGGTG 360
 GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG AGGCGGACGA CGTCCTGGCC 420
 AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC TACGAGGTCC GCATCCTCAC CGCCGACAAA 480
 GACCTTTACC AGCTCCTTTC CGACCGCATC CACGTCCTCC ACCCCGAGGG GTACCTCATC 540
 ACCCCGGCCT GGCTTTGGGA AAAGTACGGC CTGAGGCCCG ACCAGTGGGC CGACTACCGG 600
 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG GGAGAAGACG 660
 GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC TCCTCAAGAA CCTGGACCGG 720
 CTGAAGCCCG CCATCCGGGA GAAGATCCTG GCCCACATGG ACGATCTGAA GCTCTCCTGG 780
 GACCTGGCCA AGGTGCGCAC CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG 840
 CCCGACCGGG AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC 900
 GAGTTCGGCC TTCTGGAAAG CCCCTATGAC AACTACGTCA CCATCCTTGA TGAAGAAACA 960
 CTGAAAGCGT GGATTGCGAA GCTGGAAAAA GCGCCGGTAT TTGCATTTGA TACCGAAACC 1020
 GACAGCCTTG ATAACATCTC TGCTAACCTG GTCGGGCTTT CTTTGTCTAT CGAGCCAGGC 1080
 GTAGCGGCAT ATATTCCGGT TGCTCATGAT TATCTTGATG CGCCCGATCA AATCTCTCGC 1140

GAGCGTGCAC TCGAGTTGCT AAAACCGCTG CTGGAAGATG AAAAGGCGCT GAAGGTCGGG 1200
CAAAACCTGA AATACGATCG CGGTATTCTG GCGAACTACG GCATTGAACT GCGTGGGATT 1260
GCGTTTGATA CCATGCTGGA GTCCTACATT CTCAATAGCG TTGCCGGGCG TCACGATATG 1320
GACAGCCTCG CGGAACGTTG GTTGAAGCAC AAAACCATCA CTTTGAAGA GATTGCTGGT 1380
AAAGGCAAAA ATCAACTGAC CTTTAACCAG ATTGCCCTCG AAGAAGCCGG ACGTTACGCC 1440
GCCGAAGATG CAGATGTCAC CTTGCAGTTG CATCTGAAAA TGTGGCCGGA TCTGCAAAAA 1500
CACAAAGGGC CGTTGAACGT CTTGAGAAT ATCGAAATGC CGCTGGTGCC GGTGCTTTCA 1560
CGCATTGAAC GTAACGGTGT GCGCCTGGAC GTGGCCTATC TCAGGGCCTT GTCCCTGGAG 1620
GTGGCCGAGG AGATCGCCCG CCTCGAGGCC GAGGTCTTCC GCCTGGCCGG CCACCCCTTC 1680
AACCTCAACT CCCGGGACCA GCTGGAAAGG GTCCTCTTTG ACGAGCTAGG GCTTCCCGCC 1740
ATCGGCAAGA CGGAGAAGAC CGGCAAGCGC TCCACCAGCG CCGCCGTCCT GGAGGCCCTC 1800
CGCGAGGCCC ACCCCATCGT GGAGAAGATC CTGCAGTACC GGGAGCTCAC CAAGCTGAAG 1860
AGCACCTACA TTGACCCCTT GCCGGACCTC ATCCACCCCA GGACGGGCCG CCTCCACACC 1920
CGCTTCAACC AGACGGCCAC GGCCACGGGC AGGCTAAGTA GCTCCGATCC CAACCTCCAG 1980
AACATCCCCG TCCGCACCCC GCTTGGGCAG AGGATCCGCC GGGCCTTCAT CGCCGAGGAG 2040
GGGTGGCTAT TGGTGGCCCT GGACTATAGC CAGATAGAGC TCAGGGTGCT GGCCACCTC 2100
TCCGGCGACG AGAACCTGAT CCGGGTCTTC CAGGAGGGGC GGGACATCCA CACGGAGACC 2160
GCCAGCTGGA TGTTGGCGT CCCCCGGGAG GCCGTGGACC CCCTGATGCG CCGGGCGGCC 2220
AAGACCATCA ACTTCGGGGT CCTCTACGGC ATGTCGGCCC ACCGCCCTC CCAGGAGCTA 2280
GCCATCCCTT ACGAGGAGGC CCAGGCCTTC ATTGAGCGCT ACTTTCAGAG CTTCCCCAAG 2340
GTGCGGGCCT GGATTGAGAA GACCCTGGAG GAGGGCAGGA GGCGGGGGTA CGTGGAGACC 2400
CTCTTCGGCC GCCGCCGCTA CGTGCCAGAC CTAGAGGCCC GGGTGAAGAG CGTGCGGGAG 2460
GCGGCCGAGC GCATGGCCTT CAACATGCCC GTCCAGGGCA CCGCCGCCGA CCTCATGAAG 2520
CTGGCTATGG TGAAGCTCTT CCCCAGGCTG GAGGAAATGG GGGCCAGGAT GCTCCTTCAG 2580
GTCCACGACG AGCTGGTCCT CGAGGCCCCA AAAGAGAGGG CGGAGGCCGT GGCCCGGCTG 2640
GCCAAGGAGG TCATGGAGGG GGTGTATCCC CTGGCCGTGC CCCTGGAGGT GGAGGTGGGG 2700
ATAGGGGAGG ACTGGCTCTC CGCCAAGGAG TGA 2733

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2727 base pairs
- (B) Type: nucleotide
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA AATGAGGGGC   60
ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG TCGACGGCCA CCACCTGGCC  120
TACCGCACCT TCCACGCCCT GAAGGGCCTC ACCACCAGCC GGGGGGAGCC GGTGCAGGCG  180
GTCTACGGCT TCGCCAAGAG CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC  240
GTGGTCTTTG ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG  300
GGCCGGGCCC CCACGCCGGA GGA CTTTCCC CGGCAACTCG CCCTCATCAA GGAGCTGGTG  360
GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG AGGCGGACGA CGTCCTGGCC  420
AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC TACGAGGTCC GCATCCTCAC CGCCGACAAA  480
GACCTTTACC AGCTCCTTTC CGACCGCATC CACGTCCTCC ACCCGAGGG GTACCTCATC  540
ACCCCGGCCT GGCTTTGGGA AAAGTACGGC CTGAGGCCCG ACCAGTGGGC CGACTACCGG  600
GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG GGAGAAGACG  660
GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC TCCTCAAGAA CCTGGACCGG  720
CTGAAGCCCG CCATCCGGA GAAGATCCTG GCCCACATGG ACGATCTGAA GCTCTCCTGG  780
GACCTGGCCA AGGTGCGCAC CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG  840
CCCGACCGGG AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC  900
GAGTTCGGCC TTCTGGAAAG CCCCCCGTT GGATACAGAA TAGTGAAAGA CCTGGTGGAA  960
TTTGAAAAAC TCATAGAGAA ACTGAGAGAA TCCCCTTCGT TCGCCATAGA TCTTGAGACG 1020
TCTTCCCTCG ATCCTTTCGA CTGCGACATT GTCGGTATCT CTGTGTCTTT CAAACCAAAG 1080
GAAGCGTACT ACATACCACT CCATCATAGA AACGCCCAGA ACCTGGATGA AAAAGAAGTT 1140
CTGAAAAAGC TAAAAGAAAT CCTGGAGGAC CCCGGAGCAA AGATCGTTGG TCAGAATTTG 1200
AAATTCGATT ACAAGGTGTT GATGGTAAAG GGTGTTGAAC CTGTCCCTCC TCACTTCGAC 1260
ACGATGATAG CGGCTTACCT TCTTGAGCCG AACGAAAAGA AGTTCAATCT GGACGATCTC 1320
GCATTGAAAT TTCTTGATA CAAAATGACC TCTTACCAGG AACTCATGTC CTTCTCTTCT 1380
CCGCTGTTTG GTTTCAGTTT TGCCGATGTT CCTGTAGAAA AAGCAGCGAA CTATTCCTGT 1440

```

GAAGATGCCG ACATCACCTA CAGACTCTAC AAGATCCTGA GCTTAAACT CCACGAGGAG 1500
 AGGCTCCTTT GGCTTTACCG GGAGGTGGAG AGGCCCTTT CCGCTGTCCT GGGCCACATG 1560
 GAGGCCACGG GGGTGCGCCT GGACGTGGCC TATCTCAGGG CCTTGTCCCT GGAGGTGGCC 1620
 GAGGAGATCG CCCGCCTCGA GGCCGAGGTC TTCCGCCTGG CCGGCCACCC CTTCAACCTC 1680
 AACTCCCGGG ACCAGCTGGA AAGGGTCCTC TTTGACGAGC TAGGGCTTCC CGCCATCGGC 1740
 AAGACGGAGA AGACCGGCAA GCGCTCCACC AGCGCCGCCG TCCTGGAGGC CCTCCGCGAG 1800
 GCCCACCCCA TCGTGGAGAA GATCCTGCAG TACCGGGAGC TCACCAAGCT GAAGAGCACC 1860
 TACATTGACC CCTTGCCGGA CCTCATCCAC CCCAGGACGG GCCGCCTCCA CACCCGCTTC 1920
 AACCAGACGG CCACGGCCAC GGGCAGGCTA AGTAGCTCCG ATCCCAACCT CCAGAACATC 1980
 CCCGTCCGCA CCCCCTTGG GCAGAGGATC CGCCGGGCCT TCATCGCCGA GGAGGGGTGG 2040
 CTATTGGTGG CCCTGGACTA TAGCCAGATA GAGCTCAGGG TGCTGGCCCA CCTCTCCGGC 2100
 GACGAGAACC TGATCCGGGT CTTCCAGGAG GGGCGGGACA TCCACACGGA GACCGCCAGC 2160
 TGGATGTTCG GCGTCCCCCG GGAGGCCGTG GACCCCTGA TGCGCCGGGC GGCCAAGACC 2220
 ATCAACTTCG GGGTCCTCTA CGGCATGTCG GCCACCGCC TCTCCCAGGA GCTAGCCATC 2280
 CCTTACGAGG AGGCCAGGC CTTATTGAG CGCTACTTTC AGAGCTTCCC CAAGGTGCGG 2340
 GCCTGGATTG AGAAGACCCT GGAGGAGGGC AGGAGGCGGG GGTACGTGGA GACCTCTTC 2400
 GGCCGCCGCC GCTACGTGCC AGACCTAGAG GCCCGGGTGA AGAGCGTGCG GGAGGCGGCC 2460
 GAGCGCATGG CCTTCAACAT GCCCGTCCAG GGCACCGCCG CCGACCTCAT GAAGCTGGCT 2520
 ATGGTGAAGC TCTTCCCCAG GCTGGAGGAA ATGGGGGCCA GGATGCTCCT TCAGGTCCAC 2580
 GACGAGCTGG TCCTCGAGGC CCCAAAAGAG AGGGCGGAGG CCGTGGCCCG GCTGGCCAAG 2640
 GAGGTCATGG AGGGGTGTA TCCCCTGGCC GTGCCCTGG AGGTGGAGGT GGGGATAGGG 2700
 GAGGACTGGC TCTCCGCCAA GGAGTGA 2727

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2727 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA AATGAGGGGC 60

ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG TCGACGGCCA CCACCTGGCC 120
TACCGCACCT TCCACGCCCT GAAGGGCCTC ACCACCAGCC GGGGGGAGCC GGTGCAGGCG 180
GTCTACGGCT TCGCCAAGAG CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC 240
GTGGTCTTTG ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG 300
GGCCGGGCCC CCACGCCGGA GGACTTTCCC CGGCAACTCG CCCTCATCAA GGAGCTGGTG 360
GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG AGGCGGACGA CGTCCTGGCC 420
AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC TACGAGGTCC GCATCCTCAC CGCCGACAAA 480
GACCTTTACC AGCTCCTTTC CGACCGCATC CACGTCCTCC ACCCCGAGGG GTACCTCATC 540
ACCCCGGCCT GGCTTTGGGA AAAGTACGGC CTGAGGCCCG ACCAGTGGGC CGACTACCGG 600
GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG GGAGAAGACG 660
GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC TCCTCAAGAA CCTGGACCGG 720
CTGAAGCCCG CCATCCGGA GAAGATCCTG GCCCACATGG ACGATCTGAA GCTCTCCTGG 780
GACCTGGCCA AGGTGCGCAC CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG 840
CCCGACCGGG AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC 900
GAGTTCGGCC TTCTGGAAAG CCCCCCGTT GGATACAGAA TAGTGAAAGA CCTGGTGGAA 960
TTTGAAAAAC TCATAGAGAA ACTGAGAGAA TCCCCTTCGT TCGCCATAGA TCTTGAGACG 1020
TCTTCCCTCG ATCCTTTTGA CTGCGACATT GTCGGTATCT CTGTGTCTTT CAAACCAAAG 1080
GAAGCGTACT ACATACCACT CCATCATAGA AACGCCCAGA ACCTGGATGA AAAAGAAGTT 1140
CTGAAAAAGC TAAAAGAAAT CCTGGAGGAC CCCGGAGCAA AGATCGTGG TCAGAATTTG 1200
AAATTGATT ACAAGGTGTT GATGGTAAAG GGTGTTGAAC CTGTCCCTCC TCACTTCGAC 1260
ACGATGATAG CGGCTTACCT TCTTGAGCCG AACGAAAAGA AGTTCAATCT GGACGATCTC 1320
GCATTGAAAT TTCTTGATA CAAAATGACC TCTTACCAGG AACTCATGTC CTTCTCTTCT 1380
CCGCTGTTTG GTTTCAGTTT TGCCGATGTT CCTGTAGAAA AAGCAGCGAA CTATTCCTGT 1440
GAAGATGCAG ACATCACCTA CAGACTCTAC AAGATCCTGA GCTTAAAACT CCACGAGGCA 1500
GATCTGGAGA ACGTGTTCTA CAAGATAGAA ATGCCTCTTG TGAGCGTGCT TGCACGGATG 1560
GAACTGAACG GTGTGCGCCT GGACGTGGCC TATCTCAGGG CCTTGTCCCT GGAGGTGGCC 1620
GAGGAGATCG CCCGCCTCGA GGCCGAGGTC TTCCGCCTGG CCGGCCACCC CTTCAACCTC 1680
AACTCCCGGG ACCAGCTGGA AAGGGTCCTC TTTGACGAGC TAGGGCTTCC CGCCATCGGC 1740
AAGACGGAGA AGACCGGCAA GCGCTCTACC AGCGCCGCCG TCCTGGAGGC CCTCCGCGAG 1800
GCCCACCCCA TCGTGGAGAA GATCCTGCAG TACCGGGAGC TCACCAAGCT GAAGAGCACC 1860

TACATTGACC CCTTGCCGGA CCTCATCCAC CCCAGGACGG GCCGCCTCCA CACCCGCTTC 1920
 AACCAGACGG CCACGGCCAC GGGCAGGCTA AGTAGCTCCG ATCCCAACCT CCAGAACATC 1980
 CCCGTCCGCA CCCCCTTGG GCAGAGGATC CGCCGGGCCT TCATCGCCGA GGAGGGGTGG 2040
 CTATTGGTGG CCCTGGACTA TAGCCAGATA GAGCTCAGGG TGCTGGCCCA CCTCTCCGGC 2100
 GACGAGAACC TGATCCGGGT CTTCCAGGAG GGGCGGGACA TCCACACGGA GACCGCCAGC 2160
 TGGATGTTTCG GCGTCCCCCG GGAGGCCGTG GACCCCTGA TCGCGCGGGC GGCCAAGACC 2220
 ATCAACTTCG GGGTCCTCTA CGGCATGTCG GCCCACC GCC TCTCCAGGA GCTAGCCATC 2280
 CCTTACGAGG AGGCCCAGGC CTTCAATTGAG CGCTACTTTC AGAGCTTCCC CAAGGTGCGG 2340
 GCCTGGATTG AGAAGACCCT GGAGGAGGGC AGGAGGCGGG GGTACGTGGA GACCCTCTTC 2400
 GGCCGCCGCC GCTACGTGCC AGACCTAGAG GCCCGGGTGA AGAGCGTGCG GGAGGCGGCC 2460
 GAGCGCATGG CCTTCAACAT GCCCGTCCAG GGCACCGCCG CCGACCTCAT GAAGCTGGCT 2520
 ATGGTGAAGC TCTTCCCCAG GCTGGAGGAA ATGGGGGCCA GGATGCTCCT TCAGGTCCAC 2580
 GACGAGCTGG TCCTCGAGGC CCCAAAAGAG AGGGCGGAGG CCGTGGCCCG GCTGGCCAAG 2640
 GAGGTCATGG AGGGGGTGTA TCCCCTGGCC GTGCCCTGG AGGTGGAGGT GGGGATAGGG 2700
 GAGGACTGGC TCTCCGCCAA GGAGTGA 2727

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2850 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA AATGAGGGGC 60
 ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG TCGACGGCCA CCACCTGGCC 120
 TACCGCACCT TCCACGCCCT GAAGGGCCTC ACCACCAGCC GGGGGGAGCC GGTGCAGGCG 180
 GTCTACGGCT TCGCCAAGAG CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC 240
 GTGGTCTTTG ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG 300
 GGCCGGGCCC CCACGCCGGA GGA CTTTCCC CGGCAACTCG CCCTCATCAA GGAGCTGGTG 360
 GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG AGGCGGACGA CGTCCTGGCC 420
 AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC TACGAGGTCC GCATCCTCAC CGCCGACAAA 480
 GACCTTTACC AGCTCCTTTC CGACCGCATC CACGTCCTCC ACCCGAGGG GTACCTCATC 540

ACCCCGGCCT GGCTTTGGGA AAAGTACGGC CTGAGGCCCC ACCAGTGGGC CGACTACCGG 600
 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG GGAGAAGACG 660
 GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC TCCTCAAGAA CCTGGACCGG 720
 CTGAAGCCCC CCATCCGGGA GAAGATCCTG GCCCACATGG ACGATCTGAA GCTCTCCTGG 780
 GACCTGGCCA AGGTGCGCAC CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG 840
 CCCGACCGGG AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC 900
 GAGTTCGGCC TTCTGGAAAG CCCCCATCCA GCAGTTGTGG ACATCTTCGA ATACGATATT 960
 CCATTTGCAA AGAGATACCT CATCGACAAA GGCCTAATAC CAATGGAGGG GGAAGAAGAG 1020
 CTAAAGATTC TTGCCTTCGA TATAGAAACC CTCTATCACG AAGGAGAAGA GTTTGGAAAA 1080
 GGCCCAATTA TAATGATTAG TTATGCAGAT GAAAATGAAG CAAAGGTGAT TACTTGAAAA 1140
 AACATAGATC TTCCATACGT TGAGGTTGTA TCAAGCGAGA GAGAGATGAT AAAGAGATTT 1200
 CTCAGGATTA TCAGGGAGAA GGATCCTGAC ATTATAGTTA CTTATAATGG AGACTCATTC 1260
 GACTTCCCAT ATTTAGCGAA AAGGGCAGAA AACTTGGGA TTAAATTAAC CATTGGAAGA 1320
 GATGGAAGCG AGCCCAAGAT GCAGAGAATA GGCGATATGA CGGCTGTAGA AGTCAAGGGA 1380
 AGAATACATT TCGACTTGTA TCATGTAATA ACAAGGACAA TAAATCTCCC AACATACACA 1440
 CTAGAGGCTG TATATGAAGC AATTTTGGGA AAGCCAAAGG AGAAGGTATA CGCCGACGAG 1500
 ATAGCAAAAAG CCTGGGAAAG TGGAGAGAAC CTTGAGAGAG TTGCCAAATA CTCGATGGAA 1560
 GATGCAAAGG CAACTTATGA ACTCGGGAAA GAATTCCTTC CAATGGAAAT TCAGCTTTCA 1620
 GAGAGGCTCC TTTGGCTTTA CCGGGAGGTG GAGAGGCCCC TTTCCGCTGT CCTGGCCCCAC 1680
 ATGGAGGCCA CGGGGGTGCG CCTGGACGTG GCCTATCTCA GGGCCTTGTC CCTGGAGGTG 1740
 GCCGAGGAGA TCGCCCGCCT CGAGGCCGAG GTCTTCCGCC TGGCCGGCCA CCCCTTCAAC 1800
 CTCAACTCCC GGGACCAGCT GGAAAGGGTC CTCTTTGACG AGCTAGGGCT TCCCGCCATC 1860
 GGCAAGACGG AGAAGACCGG CAAGCGCTCC ACCAGCGCCG CCGTCCTGGA GGCCCTCCGC 1920
 GAGGCCACC CCATCGTGGA GAAGATCCTG CAGTACCGGG AGCTACCAA GCTGAAGAGC 1980
 ACCTACATTG ACCCCTTGCC GGACCTCATC CACCCAGGA CGGGCCGCCT CCACACCCGC 2040
 TTCAACCAGA CGGCCACGGC CACGGGCAGG CTAAGTAGCT CCGATCCCAA CCTCCAGAAC 2100
 ATCCCCGTCC GCACCCCGCT TGGGCAGAGG ATCCGCCGGG CTTTCATCGC CGAGGAGGGG 2160
 TGGCTATTGG TGGCCCTGGA CTATAGCCAG ATAGAGCTCA GGGTGCTGGC CCACCTCTCC 2220
 GGCGACGAGA ACCTGATCCG GGTCTTCCAG GAGGGGCGGG ACATCCACAC GGAGACCGCC 2280
 AGCTGGATGT TCGGCGTCCC CCGGGAGGCC GTGGACCCCC TGATGCGCCG GGCGGCCAAG 2340

ACCATCAACT TCGGGGTCCT CTACGGCATG TCGGCCCACC GCCTCTCCCA GGAGCTAGCC 2400
 ATCCCTTACG AGGAGGCCCA GGCCTTCATT GAGCGCTACT TTCAGAGCTT CCCCAGGTG 2460
 CGGGCCTGGA TTGAGAAGAC CCTGGAGGAG GGCAGGAGGC GGGGGTACGT GGAGACCCTC 2520
 TTCGGCCGCC GCCGCTACGT GCCAGACCTA GAGGCCCGGG TGAAGAGCGT GCGGGAGGCG 2580
 GCCGAGCGCA TGGCCTTCAA CATGCCCGTC CAGGGCACCG CCGCCGACCT CATGAAGCTG 2640
 GCTATGGTGA AGCTCTTCCC CAGGCTGGAG GAAATGGGGG CCAGGATGCT CCTTCAGGTC 2700
 CACGACGAGC TGGTCCTCGA GGCCCCAAAA GAGAGGGCGG AGGCCGTGGC CCGGCTGGCC 2760
 AAGGAGGTCA TGGAGGGGGT GTATCCCCTG GCCGTGCCCC TGGAGGTGGA GGTGGGGATA 2820
 GGGGAGGACT GGCTCTCCGC CAAGGAGTGA 2850

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2949 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATGAGGGGCT CGCATCACCA TCACCATCAC GCTGCTGACG ATGACGATAA AATGAGGGGC 60
 ATGCTACCGC TATTTGAGCC CAAGGGCCGG GTCCTCCTGG TCGACGGCCA CCACCTGGCC 120
 TACCGCACCT TCCACGCCCT GAAGGGCCTC ACCACCAGCC GGGGGGAGCC GGTGCAGGCG 180
 GTCTACGGCT TCGCCAAGAG CCTCCTCAAG GCCCTCAAGG AGGACGGGGA CGCGGTGATC 240
 GTGGTCTTTG ACGCCAAGGC CCCCTCCTTC CGCCACGAGG CCTACGGGGG GTACAAGGCG 300
 GGCCGGGCCC CCACGCCGGA GGA CTTTCCC CGGCAACTCG CCCTCATCAA GGAGCTGGTG 360
 GACCTCCTGG GGCTGGCGCG CCTCGAGGTC CCGGGCTACG AGGCGGACGA CGTCCTGGCC 420
 AGCCTGGCCA AGAAGGCGGA AAAGGAGGGC TACGAGGTCC GCATCCTCAC CGCCGACAAA 480
 GACCTTTACC AGCTCCTTTC CGACCGCATC CACGTCCTCC ACCCCGAGGG GTACCTCATC 540
 ACCCCGGCCT GGCTTTGGGA AAAGTACGGC CTGAGGCCCC ACCAGTGGGC CGACTACCGG 600
 GCCCTGACCG GGGACGAGTC CGACAACCTT CCCGGGGTCA AGGGCATCGG GGAGAAGACG 660
 GCGAGGAAGC TTCTGGAGGA GTGGGGGAGC CTGGAAGCCC TCCTCAAGAA CCTGGACCGG 720
 CTGAAGCCCG CCATCCGGGA GAAGATCCTG GCCCACATGG ACGATCTGAA GCTCTCCTGG 780
 GACCTGGCCA AGGTGCGCAC CGACCTGCCC CTGGAGGTGG ACTTCGCCAA AAGGCGGGAG 840

CCCGACCGGG AGAGGCTTAG GGCCTTTCTG GAGAGGCTTG AGTTTGGCAG CCTCCTCCAC 900
 GAGTTCGGCC TTCTGGAAAG CCCCGTTAGA GAACATCCAG CAGTTGTGGA CATCTTCGAA 960
 TACGATATTC CATTTGCAAA GAGATACCTC ATCGACAAAG GCCTAATACC AATGGAGGGG 1020
 GAAGAAGAGC TAAAGATTCT TGCCTTCGAT ATAGAAACCC TCTATCACGA AGGAGAAGAG 1080
 TTTGGAAAAG GCCCAATTAT AATGATTAGT TATGCAGATG AAAATGAAGC AAAGGTGATT 1140
 ACTTGGA AAA ACATAGATCT TCCATACGTT GAGGTTGTAT CAAGCGAGAG AGAGATGATA 1200
 AAGAGATTC TCAGGATTAT CAGGGAGAAG GATCCTGACA TTATAGTTAC TTATAATGGA 1260
 GACTCATTCG ACTTCCCATA TTTAGCGAAA AGGGCAGAAA AACTTGGGAT TAAATTAACC 1320
 ATTGGAAGAG ATGGAAGCGA GCCCAAGATG CAGAGAATAG GCGATATGAC GGCTGTAGAA 1380
 GTCAAGGGAA GAATACATTT C GACTTGTAT CATGTAATAA CAAGGACAAT AAATCTCCCA 1440
 ACATACACAC TAGAGGCTGT ATATGAAGCA ATTTTGGAA AGCCAAAGGA GAAGGTATAC 1500
 GCCGACGAGA TAGCAAAGC CTGGGAAAGT GGAGAGAACC TTGAGAGAGT TGCCAAATAC 1560
 TCGATGGAAG ATGCAAAGGC AACTTATGAA CTCGGGAAAG AATTCCTTCC AATGGAAATT 1620
 CAGCTTTCAA GATTAGTTGG ACAACCTTTA TGGGATGTTT CAAGGTCAAG CACAGGGAAC 1680
 CTTGTAGAGT GGTTCCTACT TAGGAAAGCC TACGAAAGAA ACGAAGTAGC TCCAAACAAG 1740
 CCAAGTGAAG AGGAGTATCA AAGAAGGCTC AGGGAGAGCT ACACAGGTGG ATTCGTGCGC 1800
 CTGGACGTGG CCTATCTCAG GGCCTTGTCC CTGGAGGTGG CCGAGGAGAT CGCCCGCCTC 1860
 GAGGCCGAGG TCTTCCGCCT GGCCGGCCAC CCCTTCAACC TCAACTCCCG GGACCAGCTG 1920
 GAAAGGGTCC TCTTTGACGA GCTAGGGCTT CCCGCCATCG GCAAGACGGA GAAGACCGGC 1980
 AAGCGCTCCA CCAGCGCCGC CGTCCTGGAG GCCCTCCGCG AGGCCACCC CATCGTGGAG 2040
 AAGATCCTGC AGTACCGGGA GCTCACCAAG CTGAAGAGCA CCTACATTGA CCCCTTGCCG 2100
 GACCTCATCC ACCCCAGGAC GGGCCGCCTC CACACCGCT TCAACCAGAC GGCCACGGCC 2160
 ACGGGCAGGC TAAGTAGCTC CGATCCCAAC CTCCAGAACA TCCCGTCCG CACCCGCTT 2220
 GGGCAGAGGA TCCGCCGGG CTTTCATCGCC GAGGAGGGGT GGCTATTGGT GGCCCTGGAC 2280
 TATAGCCAGA TAGAGCTCAG GGTGCTGGCC CACCTCTCCG GCGACGAGAA CCTGATCCGG 2340
 GTCTTCCAGG AGGGGCGGGA CATCCACACG GAGACCGCCA GCTGGATGTT CGGCGTCCCC 2400
 CGGGAGGCCG TGGACCCCT GATGCGCCGG GCGGCAAGA CCATCAACTT CGGGGTCCCTC 2460
 TACGGCATGT CGGCCACCG CCTCTCCAG GAGCTAGCCA TCCCTTACGA GGAGGCCAG 2520
 GCCTTCATTG AGCGCTACTT TCAGAGCTTC CCCAAGGTGC GGGCCTGGAT TGAGAAGACC 2580
 CTGGAGGAGG GCAGGAGGCG GGGGTACGTG GAGACCCTCT TCGGCCGCCG CCGCTACGTG 2640

CCAGACCTAG AGGCCCGGGT GAAGAGCGTG CGGGAGGCGG CCGAGCGCAT GGCCTTCAAC 2700
 ATGCCCCTCC AGGGCACCGC CGCCGACCTC ATGAAGCTGG CTATGGTGAA GCTCTTCCCC 2760
 AGGCTGGAGG AAATGGGGGC CAGGATGCTC CTTCAGGTCC ACGACGAGCT GGTCCCTCGAG 2820
 GCCCCAAAAG AGAGGGCGGA GGCCGTGGCC CGGCTGGCCA AGGAGGTCAT GGAGGGGGTG 2880
 TATCCCCTGG CCGTGCCCCT GGAGGTGGAG GTGGGGATAG GGGAGGACTG GCTCTCCGCC 2940
 AAGGAGTGA 2949

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 910 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Arg | Gly | Ser | His | His | His | His | His | His | Ala | Ala | Asp | Asp | Asp | Asp | 1 | 5 | 10 | 15 |
| Lys | Met | Arg | Gly | Met | Leu | Pro | Leu | Phe | Glu | Pro | Lys | Gly | Arg | Val | Leu | 20 | 25 | 30 | |
| Leu | Val | Asp | Gly | His | His | Leu | Ala | Tyr | Arg | Thr | Phe | His | Ala | Leu | Lys | 35 | 40 | 45 | |
| Gly | Leu | Thr | Thr | Ser | Arg | Gly | Glu | Pro | Val | Gln | Ala | Val | Tyr | Gly | Phe | 50 | 55 | 60 | |
| Ala | Lys | Ser | Leu | Leu | Lys | Ala | Leu | Lys | Glu | Asp | Gly | Asp | Ala | Val | Ile | 65 | 70 | 75 | 80 |
| Val | Val | Phe | Asp | Ala | Lys | Ala | Pro | Ser | Phe | Arg | His | Glu | Ala | Tyr | Gly | 85 | 90 | 95 | |
| Gly | Tyr | Lys | Ala | Gly | Arg | Ala | Pro | Thr | Pro | Glu | Asp | Phe | Pro | Arg | Gln | 100 | 105 | 110 | |
| Leu | Ala | Leu | Ile | Lys | Glu | Leu | Val | Asp | Leu | Leu | Gly | Leu | Ala | Arg | Leu | 115 | 120 | 125 | |
| Glu | Val | Pro | Gly | Tyr | Glu | Ala | Asp | Asp | Val | Leu | Ala | Ser | Leu | Ala | Lys | 130 | 135 | 140 | |
| Lys | Ala | Glu | Lys | Glu | Gly | Tyr | Glu | Val | Arg | Ile | Leu | Thr | Ala | Asp | Lys | 145 | 150 | 155 | 160 |
| Asp | Leu | Tyr | Gln | Leu | Leu | Ser | Asp | Arg | Ile | His | Val | Leu | His | Pro | Glu | 165 | 170 | 175 | |

09/623326

Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
 180 185 190
 Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
 195 200 205
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
 210 215 220
 Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
 225 230 235 240
 Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
 245 250 255
 Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
 260 265 270
 Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
 275 280 285
 Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 290 295 300
 Leu Glu Ser Pro Tyr Asp Asn Tyr Val Thr Ile Leu Asp Glu Glu Thr
 305 310 315 320
 Leu Lys Ala Trp Ile Ala Lys Leu Glu Lys Ala Pro Val Phe Ala Phe
 325 330 335
 Asp Thr Glu Thr Asp Ser Leu Asp Asn Ile Ser Ala Asn Leu Val Gly
 340 345 350
 Leu Ser Phe Ala Ile Glu Pro Gly Val Ala Ala Tyr Ile Pro Val Ala
 355 360 365
 His Asp Tyr Leu Asp Ala Pro Asp Gln Ile Ser Arg Glu Arg Ala Leu
 370 375 380
 Glu Leu Leu Lys Pro Leu Leu Glu Asp Glu Lys Ala Leu Lys Val Gly
 385 390 395 400
 Gln Asn Leu Lys Tyr Asp Arg Gly Ile Leu Ala Asn Tyr Gly Ile Glu
 405 410 415
 Leu Arg Gly Ile Ala Phe Asp Thr Met Leu Glu Ser Tyr Ile Leu Asn
 420 425 430
 Ser Val Ala Gly Arg His Asp Met Asp Ser Leu Ala Glu Arg Trp Leu
 435 440 445
 Lys His Lys Thr Ile Thr Phe Glu Glu Ile Ala Gly Lys Gly Lys Asn
 450 455 460
 Gln Leu Thr Phe Asn Gln Ile Ala Leu Glu Glu Ala Gly Arg Tyr Ala
 465 470 475 480
 Ala Glu Asp Ala Asp Val Thr Leu Gln Leu His Leu Lys Met Trp Pro
 485 490 495

Asp Leu Gln Lys His Glu Arg Leu Leu Trp Leu Tyr Arg Glu Val Glu
 500 505 510

Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly Val Arg
 515 520 525

Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala Glu Glu
 530 535 540

Val Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His Pro Phe
 545 550 555 560

Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu Leu
 565 570 575

Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser Thr
 580 585 590

Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val Glu
 595 600 605

Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr Tyr Ile
 610 615 620

Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu His Thr
 625 630 635 640

Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp
 645 650 655

Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg Ile
 660 665 670

Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala Leu Asp
 675 680 685

Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly Asp Glu
 690 695 700

Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr Glu Thr
 705 710 715 720

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu Met
 725 730 735

Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly Met Ser
 740 745 750

Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu Ala Gln
 755 760 765

Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg Ala Trp
 770 775 780

Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val Glu Thr
 785 790 795 800

Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg Val Lys
 805 810 815

09/623326

Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val Gln
820 825 830

Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu Phe Pro
835 840 845

Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His Asp Glu
850 855 860

Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala Arg Leu
865 870 875 880

Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro Leu Glu
885 890 895

Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
900 905 910

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 910 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Arg Gly Ser His His His His His His Ala Ala Asp Asp Asp Asp
1 5 10 15

Lys Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
20 25 30

Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
35 40 45

Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
50 55 60

Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
65 70 75 80

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
85 90 95

Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
100 105 110

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
115 120 125

Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
130 135 140

Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
 145 150 155 160
 Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
 165 170 175
 Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
 180 185 190
 Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
 195 200 205
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
 210 215 220
 Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
 225 230 235 240
 Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
 245 250 255
 Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
 260 265 270
 Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
 275 280 285
 Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 290 295 300
 Leu Glu Ser Pro Tyr Asp Asn Tyr Val Thr Ile Leu Asp Glu Glu Thr
 305 310 315 320
 Leu Lys Ala Trp Ile Ala Lys Leu Glu Lys Ala Pro Val Phe Ala Phe
 325 330 335
 Asp Thr Glu Thr Asp Ser Leu Asp Asn Ile Ser Ala Asn Leu Val Gly
 340 345 350
 Leu Ser Phe Ala Ile Glu Pro Gly Val Ala Ala Tyr Ile Pro Val Ala
 355 360 365
 His Asp Tyr Leu Asp Ala Pro Asp Gln Ile Ser Arg Glu Arg Ala Leu
 370 375 380
 Glu Leu Leu Lys Pro Leu Leu Glu Asp Glu Lys Ala Leu Lys Val Gly
 385 390 395 400
 Gln Asn Leu Lys Tyr Asp Arg Gly Ile Leu Ala Asn Tyr Gly Ile Glu
 405 410 415
 Leu Arg Gly Ile Ala Phe Asp Thr Met Leu Glu Ser Tyr Ile Leu Asn
 420 425 430
 Ser Val Ala Gly Arg His Asp Met Asp Ser Leu Ala Glu Arg Trp Leu
 435 440 445
 Lys His Lys Thr Ile Thr Phe Glu Glu Ile Ala Gly Lys Gly Lys Asn
 450 455 460

Gln Leu Thr Phe Asn Gln Ile Ala Leu Glu Glu Ala Gly Arg Tyr Ala
 465 470 475 480
 Ala Glu Asp Ala Asp Val Thr Leu Gln Leu His Leu Lys Met Trp Pro
 485 490 495
 Asp Leu Gln Lys His Lys Gly Pro Leu Asn Val Phe Glu Asn Ile Glu
 500 505 510
 Met Pro Leu Val Pro Val Leu Ser Arg Ile Glu Arg Asn Gly Val Arg
 515 520 525
 Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala Glu Glu
 530 535 540
 Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His Pro Phe
 545 550 555 560
 Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu Leu
 565 570 575
 Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser Thr
 580 585 590
 Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val Glu
 595 600 605
 Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr Tyr Ile
 610 615 620
 Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu His Thr
 625 630 635 640
 Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp
 645 650 655
 Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg Ile
 660 665 670
 Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala Leu Asp
 675 680 685
 Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly Asp Glu
 690 695 700
 Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr Glu Thr
 705 710 715 720
 Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu Met
 725 730 735
 Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly Met Ser
 740 745 750
 Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu Ala Gln
 755 760 765
 Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg Ala Trp
 770 775 780

Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val Glu Thr
 785 790 795 800
 Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg Val Lys
 805 810 815
 Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val Gln
 820 825 830
 Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu Phe Pro
 835 840 845
 Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His Asp Glu
 850 855 860
 Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala Arg Leu
 865 870 875 880
 Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro Leu Glu
 885 890 895
 Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
 900 905 910

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 908 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Met Arg Gly Ser His His His His His His Ala Ala Asp Asp Asp Asp
 1 5 10 15
 Lys Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
 20 25 30
 Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
 35 40 45
 Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
 50 55 60
 Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
 65 70 75 80
 Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 85 90 95
 Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
 100 105 110
 Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
 115 120 125

Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
 130 135 140
 Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
 145 150 155 160
 Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
 165 170 175
 Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
 180 185 190
 Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
 195 200 205
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
 210 215 220
 Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
 225 230 235 240
 Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
 245 250 255
 Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
 260 265 270
 Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
 275 280 285
 Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 290 295 300
 Leu Glu Ser Pro Pro Val Gly Tyr Arg Ile Val Lys Asp Leu Val Glu
 305 310 315 320
 Phe Glu Lys Leu Ile Glu Lys Leu Arg Glu Ser Pro Ser Phe Ala Ile
 325 330 335
 Asp Leu Glu Thr Ser Ser Leu Asp Pro Phe Asp Cys Asp Ile Val Gly
 340 345 350
 Ile Ser Val Ser Phe Lys Pro Lys Glu Ala Tyr Tyr Ile Pro Leu His
 355 360 365
 His Arg Asn Ala Gln Asn Leu Asp Glu Lys Glu Val Leu Lys Lys Leu
 370 375 380
 Lys Glu Ile Leu Glu Asp Pro Gly Ala Lys Ile Val Gly Gln Asn Leu
 385 390 395 400
 Lys Phe Asp Tyr Lys Val Leu Met Val Lys Gly Val Glu Pro Val Pro
 405 410 415
 Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Glu Pro Asn Glu
 420 425 430
 Lys Lys Phe Asn Leu Asp Asp Leu Ala Leu Lys Phe Leu Gly Tyr Lys
 435 440 445

Met Thr Ser Tyr Gln Glu Leu Met Ser Phe Ser Ser Pro Leu Phe Gly
 450 455 460
 Phe Ser Phe Ala Asp Val Pro Val Glu Lys Ala Ala Asn Tyr Ser Cys
 465 470 475 480
 Glu Asp Ala Asp Ile Thr Tyr Arg Leu Tyr Lys Ile Leu Ser Leu Lys
 485 490 495
 Leu His Glu Glu Arg Leu Leu Trp Leu Tyr Arg Glu Val Glu Arg Pro
 500 505 510
 Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly Val Arg Leu Asp
 515 520 525
 Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala Glu Glu Ile Ala
 530 535 540
 Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His Pro Phe Asn Leu
 545 550 555 560
 Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu Leu Gly Leu
 565 570 575
 Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser Thr Ser Ala
 580 585 590
 Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val Glu Lys Ile
 595 600 605
 Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr Tyr Ile Asp Pro
 610 615 620
 Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu His Thr Arg Phe
 625 630 635 640
 Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp Pro Asn
 645 650 655
 Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg Ile Arg Arg
 660 665 670
 Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala Leu Asp Tyr Ser
 675 680 685
 Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly Asp Glu Asn Leu
 690 695 700
 Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr Glu Thr Ala Ser
 705 710 715 720
 Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu Met Arg Arg
 725 730 735
 Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly Met Ser Ala His
 740 745 750
 Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu Ala Gln Ala Phe
 755 760 765

Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg Ala Trp Ile Glu
 770 775 780
 Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val Glu Thr Leu Phe
 785 790 795 800
 Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg Val Lys Ser Val
 805 810 815
 Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val Gln Gly Thr
 820 825 830
 Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu Phe Pro Arg Leu
 835 840 845
 Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His Asp Glu Leu Val
 850 855 860
 Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala Arg Leu Ala Lys
 865 870 875 880
 Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro Leu Glu Val Glu
 885 890 895
 Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
 900 905

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 908 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Arg Gly Ser His His His His His His Ala Ala Asp Asp Asp Asp
 1 5 10 15
 Lys Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
 20 25 30
 Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
 35 40 45
 Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
 50 55 60
 Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
 65 70 75 80
 Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 85 90 95
 Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
 100 105 110

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
 115 120 125
 Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
 130 135 140
 Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
 145 150 155 160
 Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
 165 170 175
 Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
 180 185 190
 Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
 195 200 205
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
 210 215 220
 Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
 225 230 235 240
 Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
 245 250 255
 Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
 260 265 270
 Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
 275 280 285
 Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 290 295 300
 Leu Glu Ser Pro Pro Val Gly Tyr Arg Ile Val Lys Asp Leu Val Glu
 305 310 315 320
 Phe Glu Lys Leu Ile Glu Lys Leu Arg Glu Ser Pro Ser Phe Ala Ile
 325 330 335
 Asp Leu Glu Thr Ser Ser Leu Asp Pro Phe Asp Cys Asp Ile Val Gly
 340 345 350
 Ile Ser Val Ser Phe Lys Pro Lys Glu Ala Tyr Tyr Ile Pro Leu His
 355 360 365
 His Arg Asn Ala Gln Asn Leu Asp Glu Lys Glu Val Leu Lys Lys Leu
 370 375 380
 Lys Glu Ile Leu Glu Asp Pro Gly Ala Lys Ile Val Gly Gln Asn Leu
 385 390 395 400
 Lys Phe Asp Tyr Lys Val Leu Met Val Lys Gly Val Glu Pro Val Pro
 405 410 415
 Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Glu Pro Asn Glu
 420 425 430

Lys Lys Phe Asn Leu Asp Asp Leu Ala Leu Lys Phe Leu Gly Tyr Lys
 435 440 445
 Met Thr Ser Tyr Gln Glu Leu Met Ser Phe Ser Ser Pro Leu Phe Gly
 450 455 460
 Phe Ser Phe Ala Asp Val Pro Val Glu Lys Ala Ala Asn Tyr Ser Cys
 465 470 475 480
 Glu Asp Ala Asp Ile Thr Tyr Arg Leu Tyr Lys Ile Leu Ser Leu Lys
 485 490 495
 Leu His Glu Ala Asp Leu Glu Asn Val Phe Tyr Lys Ile Glu Met Pro
 500 505 510
 Leu Val Ser Val Leu Ala Arg Met Glu Leu Asn Gly Val Arg Leu Asp
 515 520 525
 Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala Glu Glu Ile Ala
 530 535 540
 Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His Pro Phe Asn Leu
 545 550 555 560
 Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu Leu Gly Leu
 565 570 575
 Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser Thr Ser Ala
 580 585 590
 Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val Glu Lys Ile
 595 600 605
 Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr Tyr Ile Asp Pro
 610 615 620
 Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu His Thr Arg Phe
 625 630 635 640
 Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp Pro Asn
 645 650 655
 Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg Ile Arg Arg
 660 665 670
 Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala Leu Asp Tyr Ser
 675 680 685
 Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly Asp Glu Asn Leu
 690 695 700
 Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr Glu Thr Ala Ser
 705 710 715 720
 Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu Met Arg Arg
 725 730 735
 Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly Met Ser Ala His
 740 745 750

Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu Ala Gln Ala Phe
 755 760 765
 Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg Ala Trp Ile Glu
 770 775 780
 Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val Glu Thr Leu Phe
 785 790 795 800
 Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg Val Lys Ser Val
 805 810 815
 Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val Gln Gly Thr
 820 825 830
 Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu Phe Pro Arg Leu
 835 840 845
 Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His Asp Glu Leu Val
 850 855 860
 Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala Arg Leu Ala Lys
 865 870 875 880
 Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro Leu Glu Val Glu
 885 890 895
 Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
 900 905

(2) INFORMATION FORSEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 949 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Met Arg Gly Ser His His His His His His Ala Ala Asp Asp Asp Asp
 1 5 10 15
 Lys Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
 20 25 30
 Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
 35 40 45
 Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
 50 55 60
 Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
 65 70 75 80
 Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 85 90 95

Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
 100 105 110
 Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
 115 120 125
 Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
 130 135 140
 Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
 145 150 155 160
 Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
 165 170 175
 Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
 180 185 190
 Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
 195 200 205
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
 210 215 220
 Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
 225 230 235 240
 Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
 245 250 255
 Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
 260 265 270
 Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
 275 280 285
 Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 290 295 300
 Leu Glu Ser Pro His Pro Ala Val Val Asp Ile Phe Glu Tyr Asp Ile
 305 310 315 320
 Pro Phe Ala Lys Arg Tyr Leu Ile Asp Lys Gly Leu Ile Pro Met Glu
 325 330 335
 Gly Glu Glu Glu Leu Lys Ile Leu Ala Phe Asp Ile Glu Thr Leu Tyr
 340 345 350
 His Glu Gly Glu Glu Phe Gly Lys Gly Pro Ile Ile Met Ile Ser Tyr
 355 360 365
 Ala Asp Glu Asn Glu Ala Lys Val Ile Thr Trp Lys Asn Ile Asp Leu
 370 375 380
 Pro Tyr Val Glu Val Val Ser Ser Glu Arg Glu Met Ile Lys Arg Phe
 385 390 395 400
 Leu Arg Ile Ile Arg Glu Lys Asp Pro Asp Ile Ile Val Thr Tyr Asn
 405 410 415

006326
09/623326

Gly Asp Ser Phe Asp Phe Pro Tyr Leu Ala Lys Arg Ala Glu Lys Leu
420 425 430

Gly Ile Lys Leu Thr Ile Gly Arg Asp Gly Ser Glu Pro Lys Met Gln
435 440 445

Arg Ile Gly Asp Met Thr Ala Val Glu Val Lys Gly Arg Ile His Phe
450 455 460

Asp Leu Tyr His Val Ile Thr Arg Thr Ile Asn Leu Pro Thr Tyr Thr
465 470 475 480

Leu Glu Ala Val Tyr Glu Ala Ile Phe Gly Lys Pro Lys Glu Lys Val
485 490 495

Tyr Ala Asp Glu Ile Ala Lys Ala Trp Glu Ser Gly Glu Asn Leu Glu
500 505 510

Arg Val Ala Lys Tyr Ser Met Glu Asp Ala Lys Ala Thr Tyr Glu Leu
515 520 525

Gly Lys Glu Phe Leu Pro Met Glu Ile Gln Leu Ser Glu Arg Leu Leu
530 535 540

Trp Leu Tyr Arg Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His
545 550 555 560

Met Glu Ala Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu
565 570 575

Ser Leu Glu Val Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe
580 585 590

Arg Leu Ala Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu
595 600 605

Arg Val Leu Phe Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu
610 615 620

Lys Thr Gly Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg
625 630 635 640

Glu Ala His Pro Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr
645 650 655

Lys Leu Lys Ser Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro
660 665 670

Arg Thr Gly Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr
675 680 685

Gly Arg Leu Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg
690 695 700

Thr Pro Leu Gly Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly
705 710 715 720

Trp Leu Leu Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu
725 730 735

Ala His Leu Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly
740 745 750

Arg Asp Ile His Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg
755 760 765

Glu Ala Val Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe
770 775 780

Gly Val Leu Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala
785 790 795 800

Ile Pro Tyr Glu Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser
805 810 815

Phe Pro Lys Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg
820 825 830

Arg Arg Gly Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro
835 840 845

Asp Leu Glu Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met
850 855 860

Ala Phe Asn Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu
865 870 875 880

Ala Met Val Lys Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met
885 890 895

Leu Leu Gln Val His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg
900 905 910

Ala Glu Ala Val Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr
915 920 925

Pro Leu Ala Val Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp
930 935 940

Leu Ser Ala Lys Glu
945

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 982 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single strand
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Arg Gly Ser His His His His His Ala Ala Asp Asp Asp Asp
1 5 10 15

Lys Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
 20 25 30
 Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
 35 40 45
 Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
 50 55 60
 Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
 65 70 75 80
 Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
 85 90 95
 Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
 100 105 110
 Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
 115 120 125
 Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
 130 135 140
 Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
 145 150 155 160
 Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
 165 170 175
 Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
 180 185 190
 Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
 195 200 205
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
 210 215 220
 Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg
 225 230 235 240
 Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu
 245 250 255
 Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu
 260 265 270
 Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala
 275 280 285
 Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu
 290 295 300
 Leu Glu Ser Pro Val Arg Glu His Pro Ala Val Val Asp Ile Phe Glu
 305 310 315 320
 Tyr Asp Ile Pro Phe Ala Lys Arg Tyr Leu Ile Asp Lys Gly Leu Ile
 325 330 335

Pro Met Glu Gly Glu Glu Glu Leu Lys Ile Leu Ala Phe Asp Ile Glu
 340 345 350
 Thr Leu Tyr His Glu Gly Glu Glu Phe Gly Lys Gly Pro Ile Ile Met
 355 360 365
 Ile Ser Tyr Ala Asp Glu Asn Glu Ala Lys Val Ile Thr Trp Lys Asn
 370 375 380
 Ile Asp Leu Pro Tyr Val Glu Val Val Ser Ser Glu Arg Glu Met Ile
 385 390 395 400
 Lys Arg Phe Leu Arg Ile Ile Arg Glu Lys Asp Pro Asp Ile Ile Val
 405 410 415
 Thr Tyr Asn Gly Asp Ser Phe Asp Phe Pro Tyr Leu Ala Lys Arg Ala
 420 425 430
 Glu Lys Leu Gly Ile Lys Leu Thr Ile Gly Arg Asp Gly Ser Glu Pro
 435 440 445
 Lys Met Gln Arg Ile Gly Asp Met Thr Ala Val Glu Val Lys Gly Arg
 450 455 460
 Ile His Phe Asp Leu Tyr His Val Ile Thr Arg Thr Ile Asn Leu Pro
 465 470 475 480
 Thr Tyr Thr Leu Glu Ala Val Tyr Glu Ala Ile Phe Gly Lys Pro Lys
 485 490 495
 Glu Lys Val Tyr Ala Asp Glu Ile Ala Lys Ala Trp Glu Ser Gly Glu
 500 505 510
 Asn Leu Glu Arg Val Ala Lys Tyr Ser Met Glu Asp Ala Lys Ala Thr
 515 520 525
 Tyr Glu Leu Gly Lys Glu Phe Leu Pro Met Glu Ile Gln Leu Ser Arg
 530 535 540
 Leu Val Gly Gln Pro Leu Trp Asp Val Ser Arg Ser Ser Thr Gly Asn
 545 550 555 560
 Leu Val Glu Trp Phe Leu Leu Arg Lys Ala Tyr Glu Arg Asn Glu Val
 565 570 575
 Ala Pro Asn Lys Pro Ser Glu Glu Glu Tyr Gln Arg Arg Leu Arg Glu
 580 585 590
 Ser Tyr Thr Gly Gly Phe Val Arg Leu Asp Val Ala Tyr Leu Arg Ala
 595 600 605
 Leu Ser Leu Glu Val Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val
 610 615 620
 Phe Arg Leu Ala Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu
 625 630 635 640
 Glu Arg Val Leu Phe Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr
 645 650 655

Glu Lys Thr Gly Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu
 660 665 670
 Arg Glu Ala His Pro Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu
 675 680 685
 Thr Lys Leu Lys Ser Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His
 690 695 700
 Pro Arg Thr Gly Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala
 705 710 715 720
 Thr Gly Arg Leu Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val
 725 730 735
 Arg Thr Pro Leu Gly Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu
 740 745 750
 Gly Trp Leu Leu Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val
 755 760 765
 Leu Ala His Leu Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu
 770 775 780
 Gly Arg Asp Ile His Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro
 785 790 795 800
 Arg Glu Ala Val Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn
 805 810 815
 Phe Gly Val Leu Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu
 820 825 830
 Ala Ile Pro Tyr Glu Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln
 835 840 845
 Ser Phe Pro Lys Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly
 850 855 860
 Arg Arg Arg Gly Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val
 865 870 875 880
 Pro Asp Leu Glu Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg
 885 890 895
 Met Ala Phe Asn Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys
 900 905 910
 Leu Ala Met Val Lys Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg
 915 920 925
 Met Leu Leu Gln Val His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu
 930 935 940
 Arg Ala Glu Ala Val Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val
 945 950 955 960
 Tyr Pro Leu Ala Val Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp
 965 970 975

Trp Leu Ser Ala Lys Glu
980

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 66 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDEDNESS: single strand
 - (D) TOPOLOGY: linear
- (ii) TYPE OF MOLECULE: other nucleic acid
 - (A) DESCRIPTION: /desc = "oligonucleotide"
- (ix) CHARACTERISTIC:
 - (A) NAME/KEY: CDS
 - (B) POSITION:1..66
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

| | |
|-----------------------------------------------------------------|----|
| GAA TTC ATG AGG GGC TCG CAT CAC CAT CAC CAT CAC GCT GCT GAC GAT | 48 |
| GAC GAT AAA ATG AGG GGC | 66 |

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) TYPE OF MOLECULE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

| | | | | | | | | | | | | | | | |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Arg | Gly | Ser | His | His | His | His | His | His | Ala | Ala | Asp | Asp | Asp | Asp |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Lys Met Arg Gly | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | |